A STUDY OF CONCEPTUAL AND MATHEMATICAL KNOWLEDGE IN INTRODUCTORY MECHANICS COURSES

A Dissertation

by

MICHAEL DAVID VAN DYKE

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	William H. Bassichis
Committee Members,	Bhaskar Dutta
	Stephen A. Fulling
	Teruki Kamon
Head of Department,	George R. Welch

December 2013

Major Subject: Physics

ABSTRACT

Much of current physics education research involves the use of the Force Concept Inventory, commonly referred to as the FCI. The FCI is a conceptual inventory examination used to study student comprehension and learning of introductory mechanics. These studies often focus on comparisons between increases in performance on pre-course and post-course FCI results for two statistically significant samples; one using a traditional course structure or teaching method and the other using different techniques.

This study contains a complete statistical analysis of the FCI in order to determine its validity as a metric for measuring course success and student learning within the context of introductory mechanics courses. All the data is from students enrolled in one particular variety of Physics 218 at Texas A&M University during the Fall 2011 semester. In particular, the analysis is done for a single statistically significant sample in order to more closely examine the relationships between the FCI, mathematics skills, and student performance. It is shown that the FCI is not a valid metric for measuring student learning within an introductory physics course and that incoming mathematics skills play a critical role in student performance.

DEDICATION

In memory of my dad…

ACKNOWLEDGEMENTS

I would like to thank all the members of my committee for their suggestions and comments on the early drafts of my dissertation. Dr. Bassichis, thank you for convincing me that finishing my doctoral research would be worthwhile. It certainly has been and, without your support, I would not be here now.

To all my friends over the years, whether you suffered with me through school or helped me to avoid it, thank you for being there. I’d never have stayed focused and lucid without you.

My family has always been there for me throughout my academic career. Always having someone to talk to, especially my sister Shannon and my Mom, while I’ve been so far away has kept me going for all these years. Many thanks to all of you for your continued love and support; I love and miss all of you dearly.

Tyana, you are the love of my life and I could never have done this without you. All the late night talks about physics and statistics have been indispensible. You’ve loved and supported me through my successes and my failures. I love you, always.

NOMENCLATURE

CLES	Common Language Effect Size
FCI	Force Concept Inventory
MBT	Mechanics Baseline Test
MIE	Mathematics Inventory Examination
PER	Physics Education Research
PHYS	Physics
STEM	Science, Technology, Engineering, and Mathematics
STEP	STEM Talent Expansion Program
TAMU	Texas A&M University

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION

In recent years, physics education research (PER) has begun to play a crucial role in an increasing number of physics departments. Many different elements of physics education have been investigated, including conceptual understanding, problem-solving, use of technology, differing pedagogy, even student attitudes toward the subject. The exact details of how students learn physics are still not entirely clear. What course elements are most important and what assessment metrics can measure these course elements is of some debate. The important course elements may even differ from one course to another.

For instance, the conceptual content of a typical introductory physics course is irrelevant in future coursework for those students that will never take another mathematical science class. An emphasis on the algebraic problem solving methods may be more appropriate as the focus, in this case, because leaving the course with additional generic problem solving skills could be of greater importance than an understanding of the conceptual course material.

For engineering and physics students, however, such a course would be inappropriate. The conceptual content should be continuously relevant throughout an engineering or physics student's future coursework. This suggests that specialized introductory physics courses should be developed specifically for engineering and physics students that, in some way, increase the overall importance of the conceptual understanding of the course material itself, as well as rigorous mathematical derivations of important physical laws. Determining how students learn these concepts in all types of introductory physics courses requires a complete statistical approach to PER.

PER attempts to identify specific elements in course structure and content that directly affect student understanding and performance. Once an element has been identified, the course structure and content can be altered to better emphasize the identified course element. This research is often executed using results from various forms of pre- and post-course examinations. The average improvement for an entire set of courses using one course structure are compared to average improvement for courses using differing course structures, in an attempt to show that particular teaching methods are more effective than others.

Generally speaking, it is difficult to compare the effectiveness of courses without a common course element. There are many different varieties of introductory mechanics courses, some with dramatically different course structures and teaching methods. Pre-course and post-course inventory examinations are used primarily to measure student learning between courses by asking students a set of common questions relating to the course material. An appropriate metric would show a statistically significant increase in correlation with student performance in the course.

This work is unique because it extends beyond the usual approach of comparing gains between many statistically significant student populations that receive differing treatments. Instead, the study is comprehensive for a single sample, allowing for a more in-depth analysis. A measure of statistical effect size between different student groups within the sample is measured. In addition, a measure of statistical correlation between student performance on two different pre- and post-course examinations, the Force Concept Inventory[1] (FCI) and Mathematics Inventory Exam (MIE), and each student's common midterm examination grades is determined. The FCI is a conceptual mechanics examination commonly used in introductory physics courses and the MIE is a mathematics examination, devised specifically for this study, which includes algebra, trigonometry, and calculus. The primary goals of this study are to determine if the FCI is a valid metric for measuring student performance and to examine the effect of incoming mathematics skills on student performance in a typical introductory mechanics course.

2. BACKGROUND

2.1 The History of Physics Education Research

The research of physics teaching methods has long been considered a matter for education departments rather than physics departments. However, design of physics curricula, implementation of physics teaching methods, and statistical analysis of the effectiveness of these educational tools requires extensive knowledge of physics and significant physics teaching experience. Slowly, over the course of many decades, PER has grown in both quantity and quality. The beginning of modern PER is largely attributed to Lillian McDermott and the Physics Education Group at the University of Washington[2]. Today, PER groups can be found in most physics departments throughout the country.

The nineties was filled with major milestones in PER. The FCI, along with its counterpart the Mechanics Baseline Test[3] (MBT) was first published in 1992 and is still in wide use today as a tool for comparing teaching methods in introductory mechanics courses. In addition, the push to include PER in physics journals and meetings began in 1995 with a white paper[4] submitted to the NSF. The paper is a request to include PER as a subfield of physics that has largely shaped the way PER is viewed today.

2.2 The Force Concept Inventory and the Mechanics Baseline Test

The FCI and the MBT were simultaneously first published in 1992 as a set of inventory examinations that, together, are meant to provide a method of profiling student understanding of Newtonian mechanics. The FCI is entirely conceptual, requiring no mathematics to complete, while the MBT contains more quantitative questions that require mathematical problem-solving.

In general, the MBT has been largely overshadowed by the FCI in terms of published data, despite the intention that both examinations be used in conjunction. It has been shown[5] that there is a statistically significant correlation between performance on the MBT and the FCI. As a result, much of the research done in PER is now based upon results from the FCI alone.

Course analysis using the FCI commonly consists of a comparison between traditional lecture-style courses and courses with some kind of alternative treatment. Average performance and average gains on the FCI for each group are calculated in an attempt to show that the treatment group shows a statistically significant increase. This type of research can be insightful, but it does not highlight relationships between individual student performance on the FCI and performance on other metrics. It is necessary to analyze these individual relationships because it allows for a better understanding of the cause and effect relationship between the two variables.

2.3 Learning in Introductory Mechanics Courses

There are many different approaches taken in teaching introductory mechanics courses. The most basic difference between introductory mechanics courses, however, is between the non-calculus and calculus based varieties. These two types of courses are meant for different students; one for those in a non-science major and those in a science major. Ultimately, the focus in each course should be different because the students need to exit the course with a different set of skills.

2.3.1 Conceptual Understanding

It can be argued that Newtonian physical intuition is of greatest importance; the students will likely never encounter anything similar to the course material again, so this is the only opportunity to change the way a student thinks about the physical world. However, increased generic problem-solving skills may be of greater benefit to the students throughout their educational careers. It is not entirely clear what the primary focus should be, in this case.

In calculus based courses, especially those for engineers and physicists, it should be the course goal to instill as great an understanding of the conceptual Newtonian mechanics, as well as the underlying mathematical theory, as possible before the end of the course. The students will use the course material continuously until they graduate and will require a solid foundation if they are to continue developing their physical intuition throughout future college coursework.

2.3.2 Mathematical Problem-Solving

In non-calculus based courses, mathematical problem-solving is restricted to algebra, geometry, and trigonometry topics. Often, an introductory physics course is the last opportunity a non-calculus student will have to develop mathematical, word-based, problem-solving skills. Ensuring the students leave the course with increased mathematics problem-solving skills is essential to successfully teaching the course.

In calculus based courses, the students will have ample opportunity before they graduate to fully develop their mathematical problem-solving skills. However, mathematics is the language in which the sciences are written. Falling behind, mathematically-speaking, is both common and deadly for those in the sciences. Ensuring the students leave the course with increased mathematics problem-solving skills, especially a better understanding of the applications of calculus to the physical world, is imperative.

2.4 Failures and Misuse of Pre-Course and Post-Course Evaluation

Pre-course and post-course evaluation is a common method of comparing the benefits and pitfalls of different teaching methods within the context of the same course. What is meant by *the same course* is often unclear. Even at Texas A&M University (TAMU), there are currently *at least* four different varieties of calculus based introductory mechanics courses offered within a given semester and all of those courses are labeled as Physics (PHYS) 218. Whether these courses can be compared with each other at all is of some question since some of the courses are intended for *only* engineering students or *only* honors students, for example.

Considering the large number of publications involving the FCI, it is generally accepted that the FCI directly measures conceptual understanding of introductory mechanics concepts, despite its lack of content beyond Newton's laws. However, the truth remains unclear[6,7,8] and the FCI continues to be used in many scenarios where it may not be an appropriate metric for teaching success or student learning.

3. RESEARCH OBJECTIVES

3.1 The Validity of the FCI in Introductory Mechanics

Currently, the FCI is widely used as a metric for successful teaching in introductory mechanics, including here at TAMU. It is generally assumed that higher scores on the FCI are well correlated with student conceptual understanding and that performance in any introductory mechanics course can be evaluated under the FCI metric. In particular, the version of PHYS 218 used in this study is part of the Science, Technology, Engineering, and Mathematics (STEM) Talent Expansion Program (STEP) at TAMU. STEP is an undertaking by the engineering department at TAMU to increase the quality of education their students receive by communicating their needs in undergraduate courses outside their department, such as physics or mathematics. This study will determine the validity of the FCI within the context of PHYS 218 STEP at TAMU and closely examine the relationship between performance on the FCI and performance in introductory mechanics courses.

3.2 The Importance of Mathematics in Introductory Mechanics

It is clear that student understanding of mathematics, particularly calculus, must play a critical role in introductory mechanics courses. After all, it is the language in which Newtonian mechanics is written and, without it, consistent communication of Newtonian concepts would be impossible. In addition, the predictive power of physics is only useful and well understood because of the underlying mathematical structure. This study will examine the relationship between student understanding of mathematics and performance in introductory mechanics courses.

4. EXPERIMENTAL PROCEDURE

4.1 Introductory Mechanics Course Selection and Structure

PHYS 218 at TAMU is a calculus-based, introductory mechanics course. In general, students enrolled in the course are majoring in a science-related field, such as biology, chemistry, engineering, pre-med, and physics. However, depending on the intended field of science, these courses should have a differing emphasis that better encompasses the post-course relevance of the course material, such as more rigorous use of the complete calculus definitions.

PHYS 218 STEP courses are designed specifically for engineering majors with an emphasis on more mathematically rigorous definitions and applications. A concrete understanding, both conceptual and illustrative in nature, of basic calculus is required throughout the course. All aspects of the course, including the textbook[9], homework, and midterm exams consistently enforce an understanding of the complete calculus-based definitions.

Lecture is traditional in nature, consisting of theoretical proofs and explanations, supported by worked-out example problems demonstrating how the theory is used to model real applications. The theory requires substantial knowledge of calculus to fully understand and reinforces the mathematical definitions of classical mechanics.

During recitation, the students take a quiz consisting of traditional work-out problems related to current topics covered in lecture. The teaching assistant answers questions and assists students in solving problems from the textbook and from previous years' exams. These activities encourage students to develop independent problem-solving skills.

The time in laboratory is spent working on physics-related, team-building exercises that provide students with real-life scenarios in which they are meant to use the material they have learned in the course to solve problems in a group environment. This includes both traditional laboratory experiments and group problem-solving exercises. Encouraging group interaction helps prepare future engineers for team projects and work environments that are commonplace in all fields of science and engineering.

PHYS 218 STEP was chosen as the focus of this study for multiple reasons. Enrolled students use the same textbook, have very similar classroom experiences, and take the same midterm examinations. In addition, the midterm exams are all graded using the same grading rubric by the same group of people. If these items were different for each student, it could lead to false conclusions, so PHYS 218 STEP was chosen to eliminate these potential problems.

4.2 Common Midterm Examinations

Three common midterm exams were given to all the students throughout the semester. The first exam covers kinematics and early applications of Newton's laws; the second covers more comprehensive applications of Newton's laws, the work-energy theorem, and potential functions; the third covers polar coordinates, Newton's laws using polar coordinates, momentum conservation, torque, and angular momentum conservation. All students are given the exams simultaneously, to prevent possible cheating, and are given the same amount of time to complete each exam. All the topics covered on these exams are commonly found in traditional introductory mechanics courses.

4.3 Pre-Course and Post-Course Evaluation

Each student was given the FCI at the beginning and the end of the course. The FCI was chosen as the tool to measure conceptual understanding here, primarily because it is used in a significant number of publications and its validity as a metric is assumed, but unclear.

A mathematics inventory examination was devised that, not only requires the students to perform traditional algebra, trigonometry, and calculus problems, but also contains questions requiring a written response about the nature of calculus. Each student was given the entire MIE at the beginning of the course and only the calculus portion of the MIE at the end of the course.

4.4 Collection of Student Data

All student data was collected at the end of the semester and can be found in the appendix. All responses to all questions on the pre-course and post-course FCI and MIE are recorded for each student, along with their score on all three midterm examinations. In addition, each student’s name and section number has been replaced with a random number to ensure anonymity. Much of the analysis is individual in nature, so only students with a complete data set were used throughout the data analysis.

5. DATA ANALYSIS AND RESULTS

5.1 Mathematical Background

There are three types of statistical analysis used in this study: the two-tailed t-Test[10], statistical effect size, and regression analysis. The purpose of a t-Test, in general, is to determine the probability that two measured sample means are actually from populations with the same mean and the observed difference is entirely due to sampling variance. In this study, a Welch's t-Test[11] will be used, so the population will be assumed Gaussian and the sample variances are assumed to be unequal. Measures of effect size attempt to quantify the magnitude of the relationship between populations or between paired data. A regression analysis measures correlation between two variables to determine the fraction of variance in the two variables that is due to their mutual relationship and how much is unexplained. The underlying statistical mathematics is relatively straight forward.

5.1.1 Welch's t-Test Analysis

Suppose that there exists a Gaussian population with a population mean μ and a population variance σ^2. A single sample of size n is taken from the population with the following sample statistics:

$$x_j = \{x_1, x_2, \ldots, x_n\} \tag{5.1}$$

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j \tag{5.2}$$

$$s^2 = \frac{1}{\nu}\sum_{j=1}^{n}\left(x_j - \bar{x}\right)^2 \tag{5.3}$$

Here, the x_j are the sampled data from the Gaussian population, $\bar{x}$ is the sample mean, s^2 is the sample variance, and $\nu = n - 1$ is the sample degrees of freedom. A more convenient set of values are the corresponding standardized normal quantities of the x_j:

$$z_j = \frac{x_j - \mu}{\sigma} = \frac{x_j - \bar{x}}{s} \tag{5.4}$$

The best estimators for the true population mean and variance are simply the sample mean and variance, so they are used here. The z_j are each a measure of the number of sample standard deviations the corresponding x_j is from the sample mean and are said to be z-distributed. The z-distribution is simply a Gaussian distribution with a mean of zero and a variance of one.

When more than one sample is taken, one creates what is referred to as a *sampling distribution*. Suppose that k samples are taken each with separate sample means $\bar{x}_i$, sample variances s_i^2, and sample sizes n_i. The mean and variance of the sample means are given by the weighted sums:

$$\hat{\mu} = \sum_{i=1}^{k} n_i \bar{x}_i \tag{5.5}$$

$$\hat{\sigma}^2 = \sum_{i=1}^{k} \frac{s_i^2}{n_i} \tag{5.6}$$

The sample means are always t-distributed, shown in Figure 1, with a mean $\hat{\mu}$ and variance $\hat{\sigma}^2$. However, unlike Gaussian distributions, t-distributions are dependent upon the degrees of freedom associated with the sample variance. Since each sample in the sampling distribution has a different sample size, the degrees of freedom must be approximated. The Welch-Satterthwaite equation[12] is used to approximate the degrees of freedom when the sample variances and sizes cannot be assumed equal:

$$\nu \approx \left(\sum_{i=1}^{k} \frac{s_i^2}{n_i}\right)^2 \Bigg/ \left(\sum_{i=1}^{k} \frac{\{s_i^2/n_i\}^2}{\nu_i}\right) \tag{5.7}$$

In a similar fashion to that of Equation (5.4), each sample mean in the sampling distribution can be standardized into a value that measures its distance from $\hat{\mu}$ in terms of $\hat{\sigma}$, referred to as its t-value:

$$t_i = \frac{\bar{x}_i - \hat{\mu}}{\hat{\sigma}} \tag{5.8}$$

The t_i are t-distributed with a mean of zero, a variance of $\frac{\nu}{\nu-2}$, and degrees of freedom approximated by Equation (5.7).

In addition to the t_i, the z_j contained in each sample can be squared and summed to create a χ^2-distributed variable, seen in Figure 2, for each sample:

$$\chi_i^2 = \left\{\sum_{j=1}^{n} z_j^2\right\}_i = \left\{\sum_{j=1}^{n} \left(\frac{x_j - \mu}{\sigma}\right)^2\right\}_i = \left\{\sum_{j=1}^{n} \left(\frac{x_j - \hat{\mu}}{\hat{\sigma}}\right)^2\right\}_i$$

$$= \left\{\frac{1}{\hat{\sigma}^2} \sum_{j=1}^{n} \left(x_j - \hat{\mu}\right)^2\right\}_i = \left\{\frac{\nu s^2}{\hat{\sigma}^2}\right\}_i = \frac{\nu_i s_i^2}{\hat{\sigma}^2} = \frac{(n_i - 1)s_i^2}{\hat{\sigma}^2} \tag{5.9}$$

The χ^2-distribution is dependent upon the degrees of freedom, with a mean of ν and a variance of 2ν. Again, the degrees of freedom of the sampling distribution must be

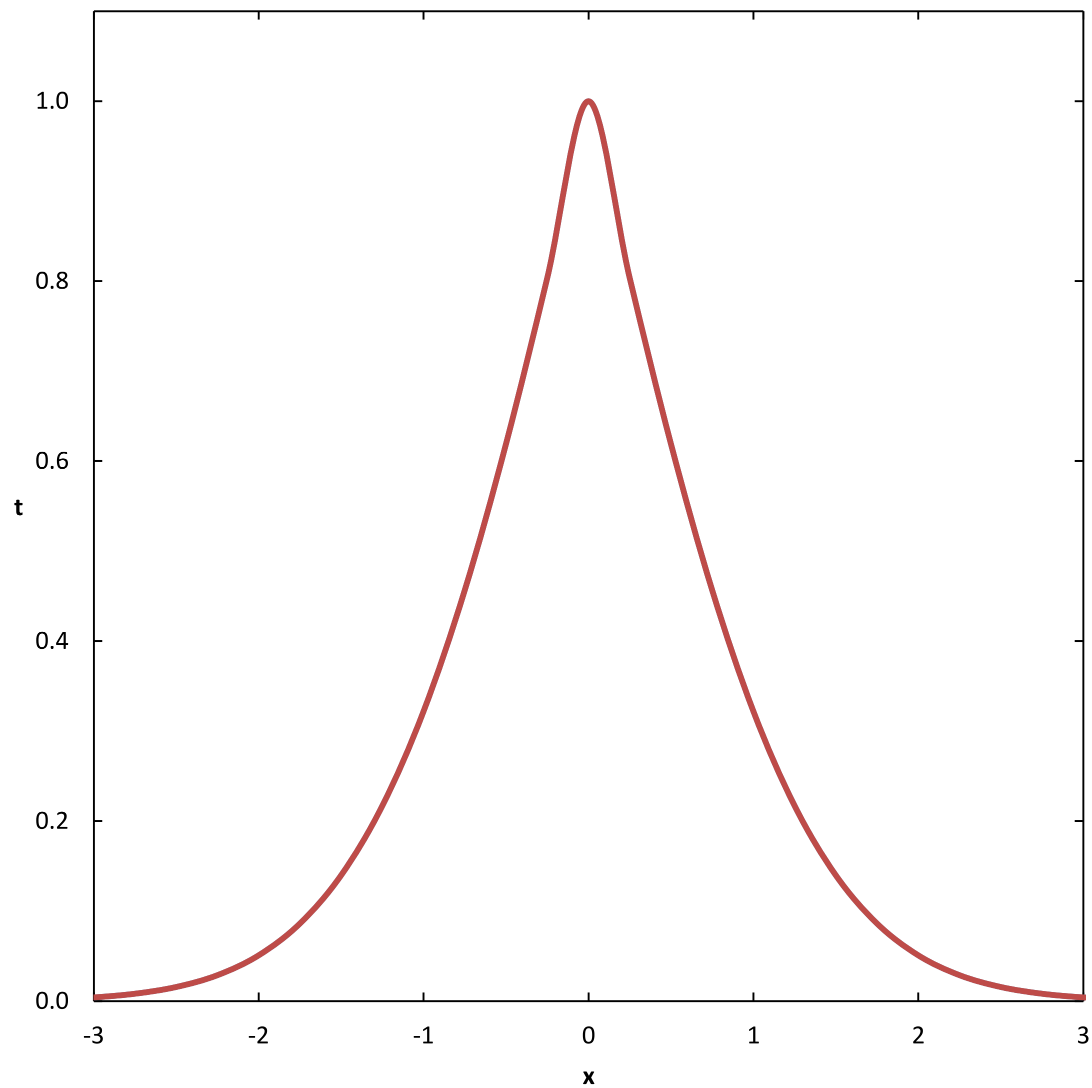

Figure 1. *Student's t Distribution.* For illustrative purposes, $\nu = 50$.

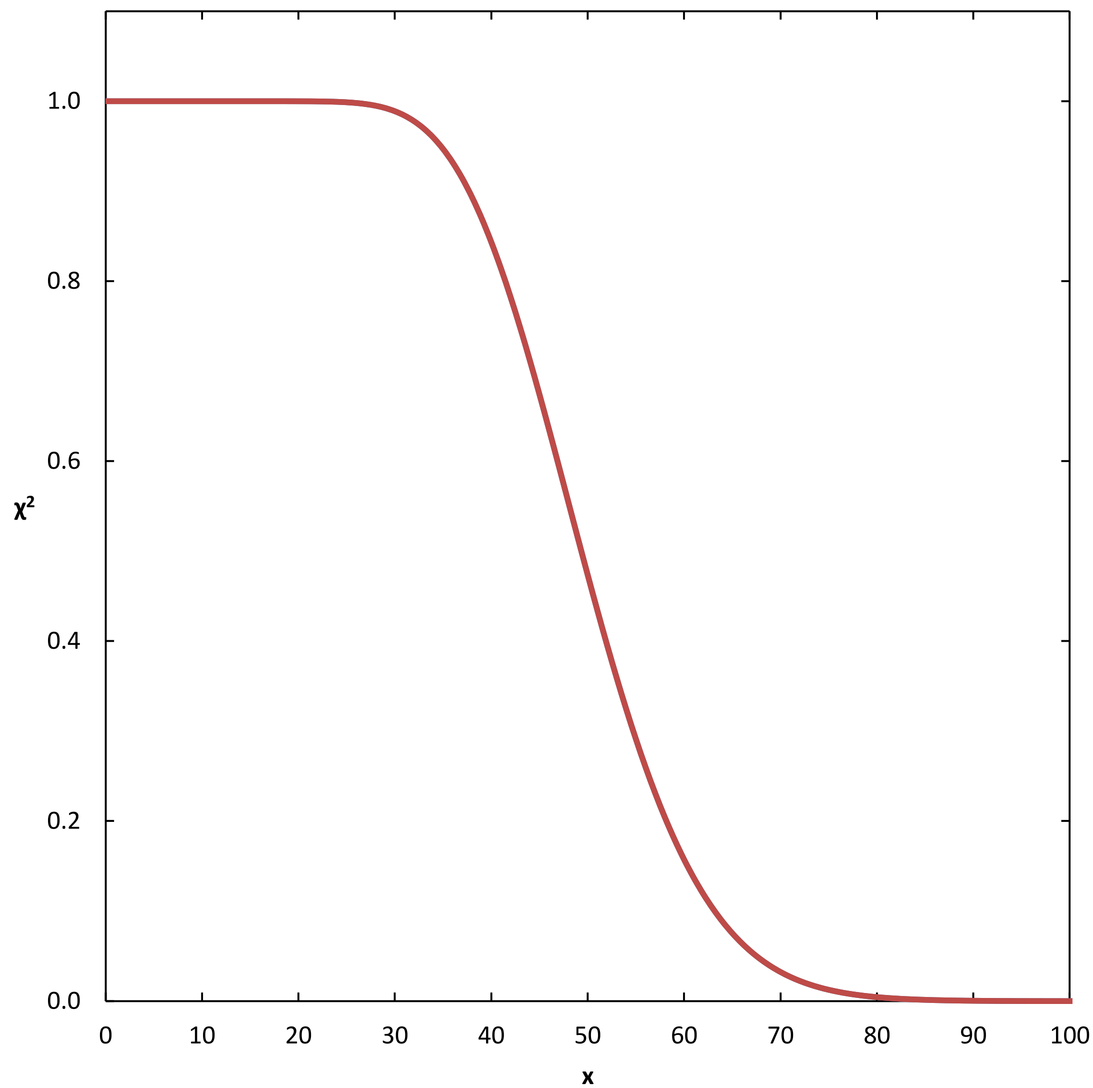

Figure 2. χ^2 *Distribution.* For illustrative purposes, $\nu = 50$.

approximated according to Equation (5.7) since the sample variances and sizes are not equal.

In the case of a Welch's t-Test, there are many samples and the intent is to measure the probability that any two samples come from the same population. Equation (5.8) then becomes:

$$t_{ij} = \frac{\bar{x}_{ij} - \mu_{ij}}{\hat{\sigma}} = \frac{\bar{x}_{ij}}{\hat{\sigma}} = \frac{\bar{x}_{ij}}{\sqrt{\sum \frac{s_i^2}{n_i}}} \tag{5.10}$$

$$where\ \bar{x}_{ij} = \bar{x}_i - \bar{x}_j\ and\ \mu_{ij} = \mu_i - \mu_j = 0$$

When there are exactly two samples in the sampling distribution, Equation (5.10) and (5.7) become:

$$t_{21} = -t_{12} = \frac{\bar{x}_2 - \bar{x}_1}{\sqrt{\frac{s_2^2}{n_2} + \frac{s_1^2}{n_1}}} \tag{5.11}$$

$$\nu \approx \left(\frac{s_2^2}{n_2} + \frac{s_1^2}{n_1}\right)^2 \Big/ \left(\frac{s_2^4}{n_2^2(n_2 - 1)} + \frac{s_1^4}{n_1^2(n_1 - 1)}\right) \tag{5.12}$$

The t-value statistics from Equation (5.8) or (5.10), along with the degrees of freedom from Equation (5.7), can be converted into a probability from the t-distribution. Its corresponding cumulative distribution function, the integral of its probability density function, will yield the probability of interest and is given by:

$$p = \frac{\Gamma\left(\frac{\nu + 1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \int_{-\infty}^{t} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2} dx = 1 - \frac{1}{2} I\left(\frac{\nu}{t^2 + \nu}; \frac{\nu}{2}, \frac{1}{2}\right) \tag{5.13}$$

$$where\ I(x;a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^x y^{a-1}(1-y)^{b-1}dy$$

Here, Γ is the gamma function and I is the regularized beta function. Equation (5.13), referred to as the p-value, is the probability that the mean difference $\bar{x}_{ij}$ was observed when the actual population mean difference is zero. The occurrence of this scenario is considered the null hypothesis. Generally, a statistical confidence level is chosen as a cutoff for what probability is necessary to reject the null hypothesis. Equation (5.11), (5.12), and (5.13) will be used throughout to calculate statistical reliability.

In addition to calculations of statistical reliability, one must also determine the uncertainty in such calculations. A *confidence interval* provides a way to consistently report the statistical uncertainty inherent in a calculation based on randomly sampled data. Equation (5.8) and (5.9), with $k = 1$, yield the corresponding confidence intervals for the mean and the variance of an individual sample, respectively:

$$\bar{x} + t_{n-1,\frac{\alpha}{2}}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{n-1,1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}} \tag{5.14}$$

$$\frac{(n-1)s^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{n-1,\frac{\alpha}{2}}} \tag{5.15}$$

These confidence intervals correspond to a $1 - \alpha$ probability that the true population parameter lies within the bounds. A commonly chosen confidence interval is 95%, meaning that there is a 5% probability, corresponding to $\alpha = .05$, that the Gaussian population parameter estimated by the sample statistic lies outside the calculated interval and this standard will be observed here. To calculate the confidence intervals in

Equation (5.14) and (5.15), it is necessary to determine the bounds of the t-distribution and χ^2-distribution that contain $1 - \alpha$ of the area under the probability density functions. More specifically, the integral of the probability density function for each distribution is set equal to the probability that the population parameter is found within the bounds of the integral:

$$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \int_{-\infty}^{t_{\alpha/2}} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2} dx = \frac{\alpha}{2} \tag{5.16}$$

$$\frac{1}{2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right)} \int_{0}^{\chi^2_{\alpha/2}} x^{(\nu/2)-1} e^{-x/2} dx = \frac{\alpha}{2} \tag{5.17}$$

Evaluating the integrals, as well as the corresponding integrals for $1 - \frac{\alpha}{2}$, and solving for the upper-bound variable allows for calculation of the confidence intervals of Equation (5.14) and (5.15). Evaluation of the integrals here is often tedious. In practice, for large sample sizes, it is useful to know that $t_{\nu,p} \to z_p$ and $\frac{\nu}{\chi^2_{\nu,p}} \to 1 - z_p\sqrt{\frac{2}{n}}$, due to the asymptotic properties of both distributions and that $z_p = z_{1-p}$ due to the symmetry about the origin for a standard normal Gaussian distribution. Under this limit, Equation (5.14) and (5.15) become:

$$\bar{x} - |z_{\alpha/2}|\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + |z_{\alpha/2}|\frac{s}{\sqrt{n}} \tag{5.18}$$

$$s^2 - |z_{\alpha/2}|\sqrt{\frac{2}{n}}s^2 \le \sigma^2 \le s^2 + |z_{\alpha/2}|\sqrt{\frac{2}{n}}s^2 \tag{5.19}$$

The large sample size limit will be used throughout to calculate confidence intervals for all sample means and variances. A 95% confidence interval will be reported with all sample means and variances. All other reported quantities will include confidence intervals constructed using the upper and lower bounds for μ and σ.

5.1.2 *Effect Size*

A measure of effect size is necessary to fully complement the p-values in a Welch's t-Test. In the context of a t-Test, an effect size is simply a standardized measure of the difference between the mean of a control sample and the mean of a treatment sample. In particular, Cohen's effect size[13] is widely used for Gaussian populations and is given by:

$$d = \frac{\bar{X}_2 - \bar{X}_1}{\sqrt{(\nu_2 s_2^2 + \nu_1 s_1^2)/(\nu_2 + \nu_1)}} \tag{5.20}$$

This quantity is the difference between the two sample means, divided by the *pooled standard deviation* which is simply the square root of an average of the two sampled variances, weighted by their degrees of freedom. It is a measure of the number of standard deviations between the two sample means. As widely used as Cohen's effect size may be, it is unfortunate that it can take values from zero to infinity. This leads to arbitrary choices for what is deemed a *large* or *small* effect size.

A *common language effect size*, or CLES[14], is a recent alternative to reporting traditional effect size. Simply stated, a CLES is the probability that a randomly chosen data point sampled from the treatment population will be larger than a randomly chosen data point sampled from the control population. It is assumed that both populations are Gaussian, so we use Cohen's effect size within the integral of the normal Gaussian probability density function:

$$CLES = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d} e^{-x^2/2} dx \tag{5.21}$$

5.1.3 Regression Analysis

Regression analysis measures correlation between two paired variables within the same sample. Traditionally[15], correlation coefficients for each sample are calculated and reported and this will be observed here. The correlation coefficient is the covariance of the two variables divided by the product of their standard deviations:

$$r = \left[\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})\right] \Bigg/ \sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2 \sum_{i=1}^{N} (y_i - \bar{y})^2} \tag{5.22}$$

Here, N is the number of data pairs, the x_i and y_i are the paired sample data, and $\bar{x}$ and $\bar{y}$ are the paired sample data means.The correlation coefficient is closely related to another important regression analysis quantity referred to as the coefficient of determination. The coefficient of determination is a measure of the correlated variance between two paired variables and is defined as:

$$r^2 = \frac{SS_{reg}}{SS_{tot}} = \left[\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})\right]^2 \Bigg/ \left[\sum_{i=1}^{N} (x_i - \bar{x})^2 \sum_{i=1}^{N} (y_i - \bar{y})^2\right] \tag{5.23}$$

This quantity is an effect size for regression analysis because it measures the percentage of the variance in the paired data that is shared between them. The remaining variance is not accounted for simply via the relationship between the two variables and said to be unexplained variance.

A measure of statistical reliability is also necessary when discussing correlation. Correlation coefficients for many samples are a t-distributed quantity, from Equation (5.8):

$$t_i = \frac{r_i - R}{\hat{\sigma}} \tag{5.24}$$

Here, the r_i are the sample correlation coefficients and R is the true population correlation coefficient. Treating the paired sample as a single sample for the purposes of calculating the t-value means that s^2 is the unexplained variance between the paired samples. Since three additional statistics of the paired sample are known (the two sample means and the correlation coefficient), the degrees of freedom for the paired sample is given by $\nu = N - 3$. This gives an *effective sample size* of $n = N - 2$. In this case, the probability that r was measured from the sampled data, when R is actually zero, is the measurement of interest, so $R = 0$, $s^2 = 1 - r^2$, $n = N - 2$, and $\hat{\sigma} = \sum \frac{s_i^2}{n_i} = \frac{1-r^2}{N-2}$. Equation (5.23) then evaluates to:

$$t = r\sqrt{\frac{N-2}{1-r^2}} \tag{5.25}$$

This t-value, along with the degrees of freedom $\nu = N - 3$, can be used in Equation (5.13) to calculate a corresponding p-value of statistical confidence.

Confidence intervals for the correlation coefficient of a sample are constructed in the same manner as for the sample mean in Equation (5.14). Using the same approximation for large sample size gives the following confidence interval:

$$r - |z_{\alpha/2}|\sqrt{\frac{1-r^2}{N-2}} \leq R \leq r + |z_{\alpha/2}|\sqrt{\frac{1-r^2}{N-2}} \tag{5.26}$$

A 95% confidence interval will be reported with all sample correlation coefficients. All other reported quantities will include confidence intervals constructed using the upper and lower bounds for r.

5.2 Incoming Student Distributions

The incoming student distribution is relatively typical for an introductory mechanics course. From a mathematics perspective, most students enter the course with a sufficient general mathematics understanding, as shown in Figure 3. More specifically, student pre-calculus understanding is very high while there seems to be a large spread in calculus knowledge among the students upon entering the course, as shown in Figures 4 and 5. This is understandable considering the first calculus course can be taken concurrently with PHYS 218 at TAMU, despite the inherent calculus-based nature of the course. Performance on the pre-course FCI is also relatively typical for an introductory mechanics course, as seen in Figure 6. The two comparative distributions are included to illustrate typical incoming student performance on the FCI for traditional lecture-style courses with similar pre-requisites (Hake) and courses that require two semesters of calculus as pre-requisites (Caballero et al.)[16].

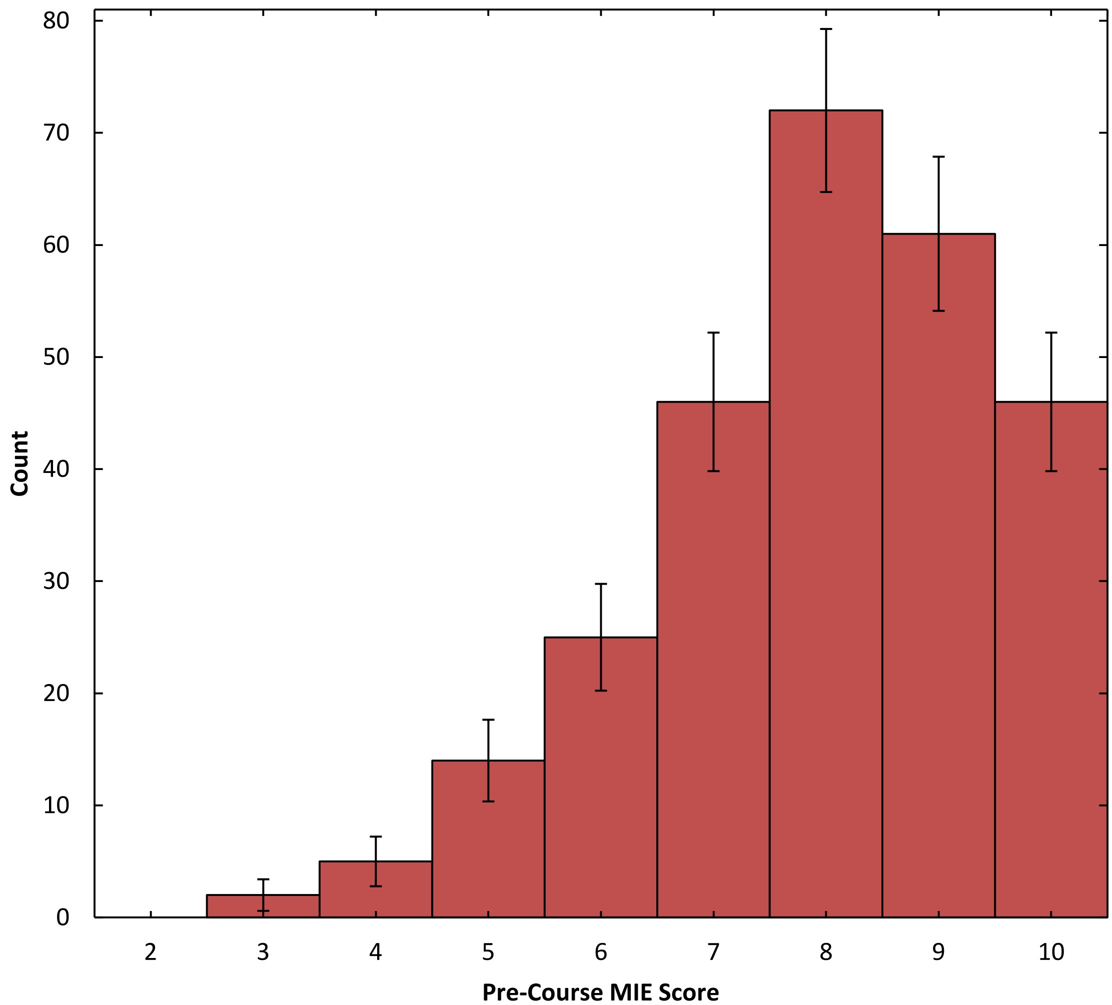

Figure 3. *Pre-Course MIE Distribution*. The distribution of student performance on the pre-course MIE. The sample statistics for this distribution are shown in Table 1.

Table 1. *Pre-Course MIE Distribution Statistics*. The sample statistics for the pre-course MIE for this study.

Sample Size (n)	***Sample Mean*** $(\bar{x})$	***Sample Variance*** (s^2)
271	7.94 ± 0.18	2.39 ± 0.40

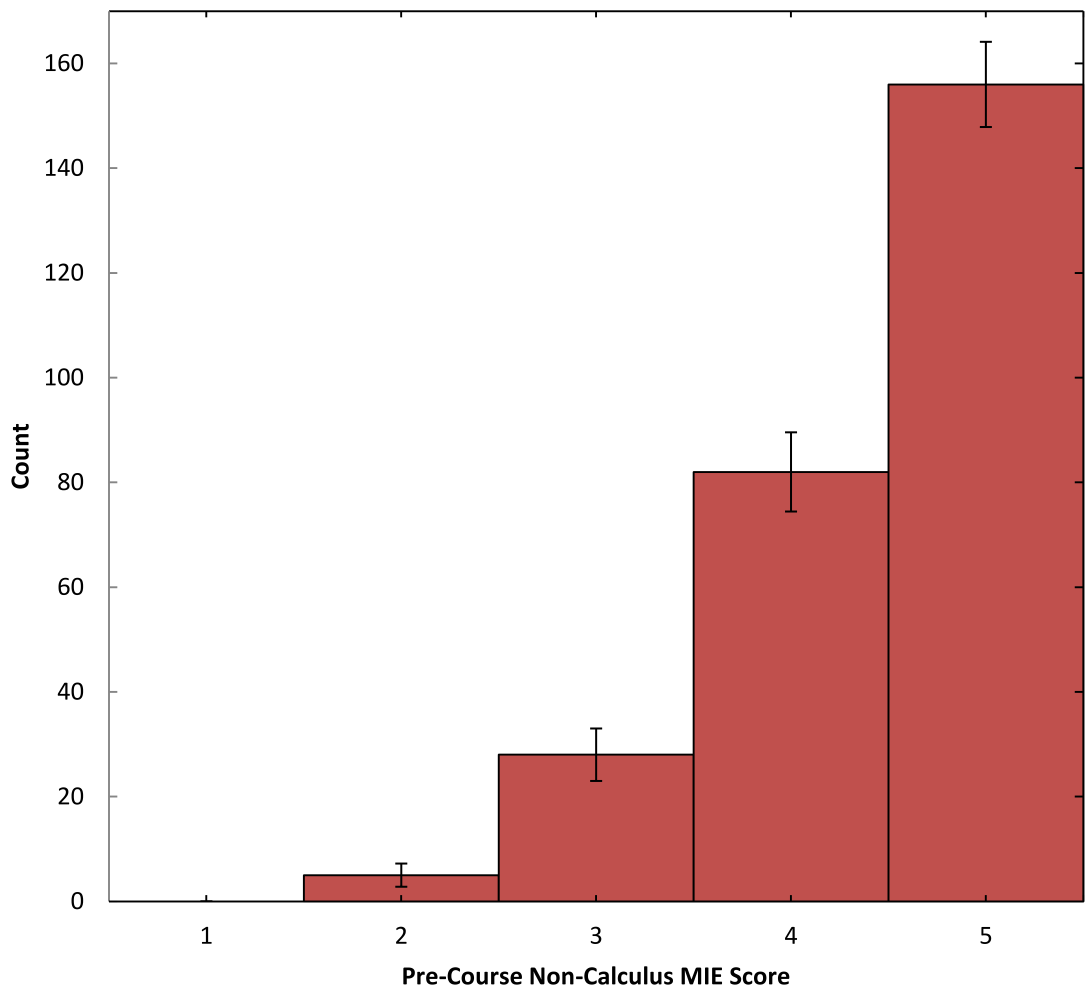

Figure 4. *Pre-Course Non-Calculus MIE Distribution*. The distribution of student performance on the non-calculus questions of the pre-course MIE. The sample statistics for this distribution are shown in Table 2.

Table 2. *Pre-Course Non-Calculus MIE Distribution Statistics*. The sample statistics for the non-calculus questions of the pre-course MIE for this study.

Sample Size (n)	*Sample Mean* $(\bar{x})$	*Sample Variance* (s^2)
271	4.44 ± 0.09	0.57 ± 0.10

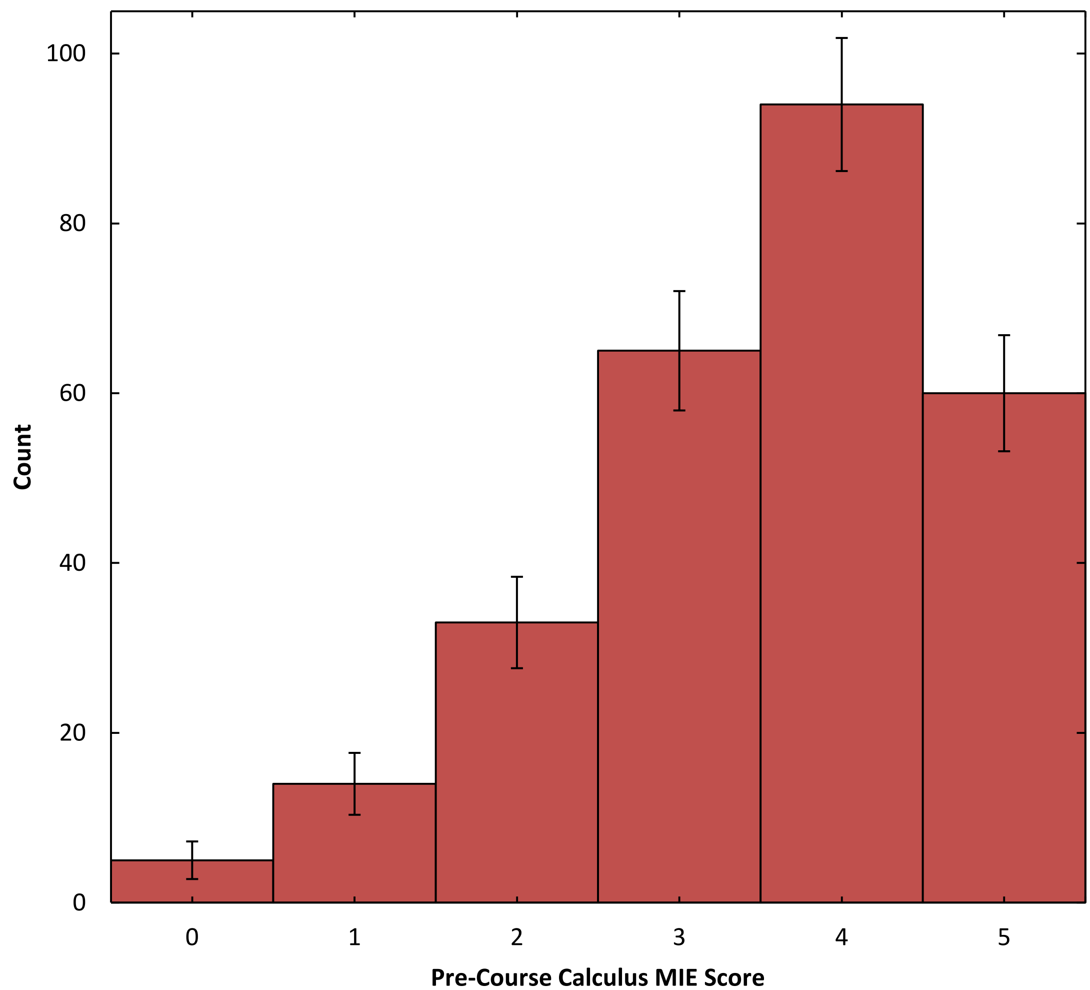

Figure 5. *Pre-Course Calculus MIE Distribution.* The distribution of student performance on the calculus questions of the pre-course MIE. The sample statistics for this distribution are shown in Table 3.

Table 3. *Pre-Course Calculus MIE Distribution Statistics.* The sample statistics for the calculus questions of the pre-course MIE for this study.

Sample Size (n)	***Sample Mean*** $(\bar{x})$	***Sample Variance*** (s^2)
271	3.51 ± 0.14	1.47 ± 0.25

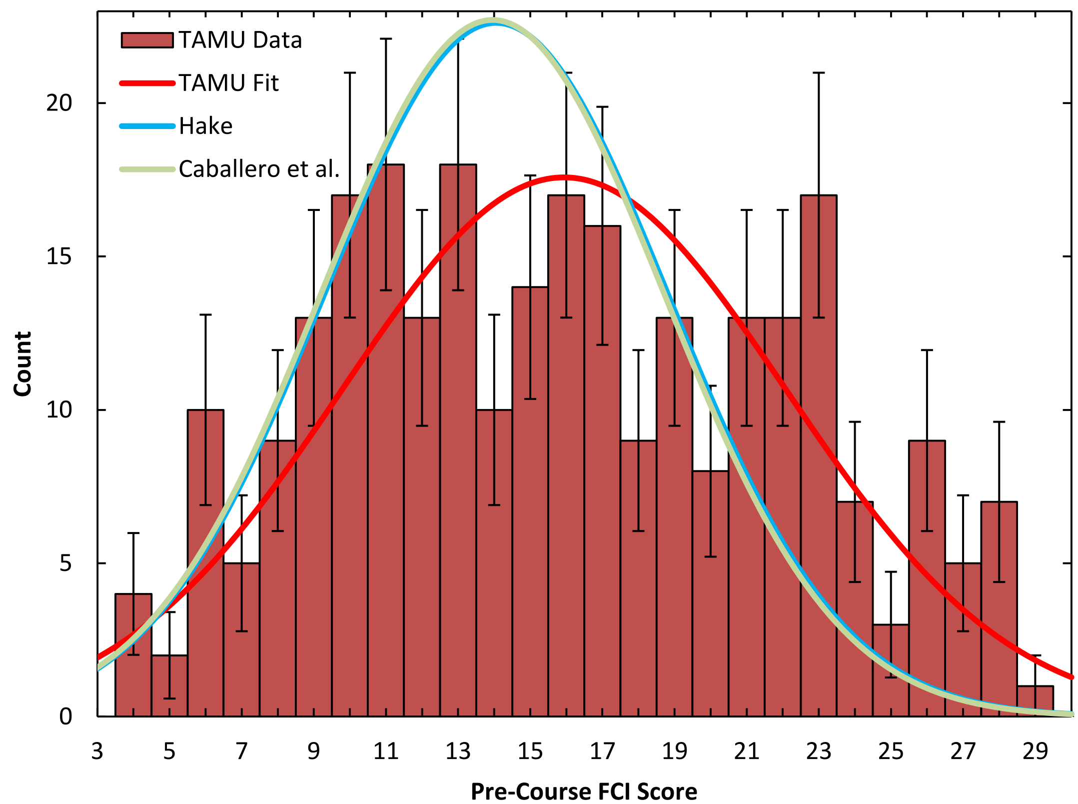

Figure 6. *Pre-Course FCI Distribution*. A comparison between data obtained for this study and two other pre-course distributions (Hake, Caballero et al.). The non-TAMU fits are scaled to a sample size equivalent to the TAMU fit. The sample statistics for this distribution are shown in Table 4.

Table 4. *Pre-Course FCI Distribution Statistics.* The sample statistics for the pre-course FCI for this study and two other distributions (Hake, Caballero et al.).

Data Set	*Sample Size* (n)	*Sample Mean* $(\bar{x})$	*Sample Variance* (s^2)
TAMU	271	15.93 ± 0.73	37.84 ± 6.37
Hake	975	14.05 ± 0.30	22.86 ± 2.03
Caballero et al.	2947	13.96 ± 0.17	22.66 ± 1.16

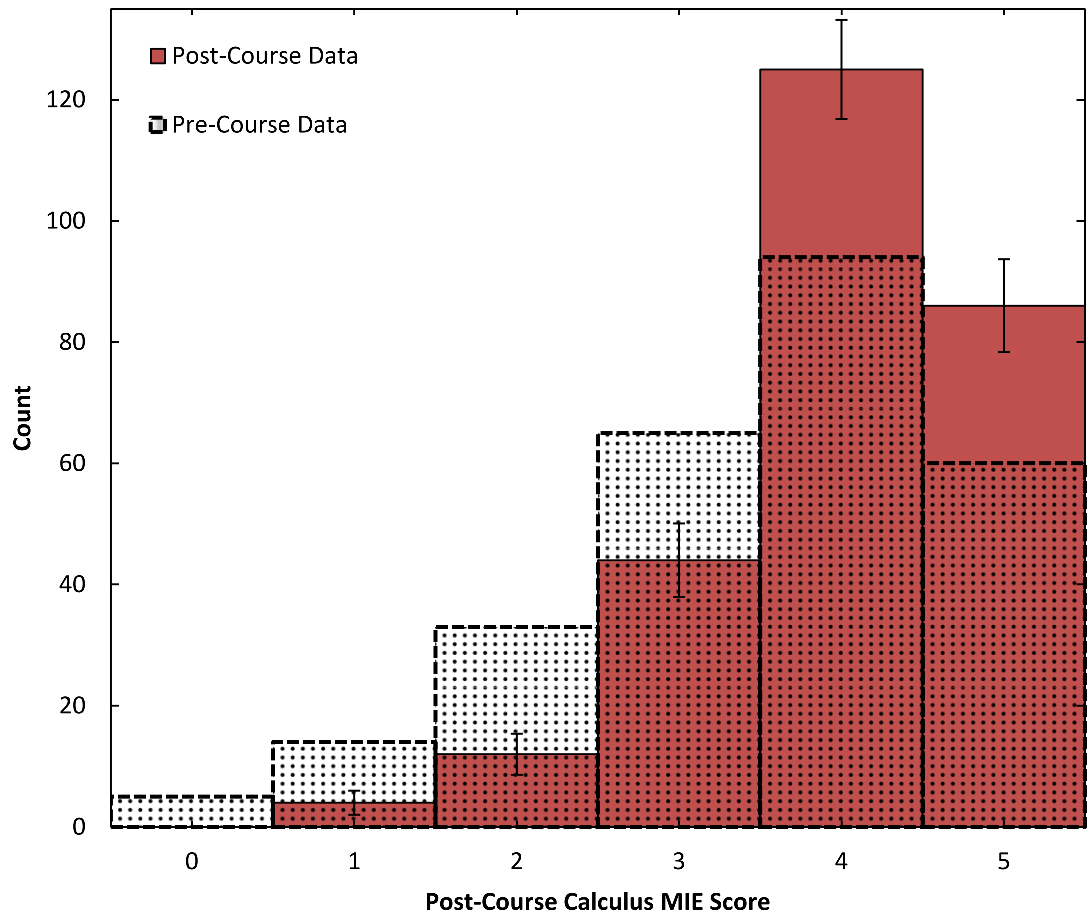

Figure 7. *Post-Course Calculus MIE Distribution*. The distribution of student performance on the calculus questions of the post-course MIE, with an overlay of the pre-course calculus MIE distribution. The sample statistics for this distribution are shown in Table 5.

Table 5. *Post-Course Calculus MIE Distribution Statistics*. The sample statistics for the calculus questions of the post-course MIE for this study.

Sample Size (n)	***Sample Mean*** $(\bar{x})$	***Sample Variance*** (s^2)
271	4.02 ± 0.11	0.79 ± 0.13

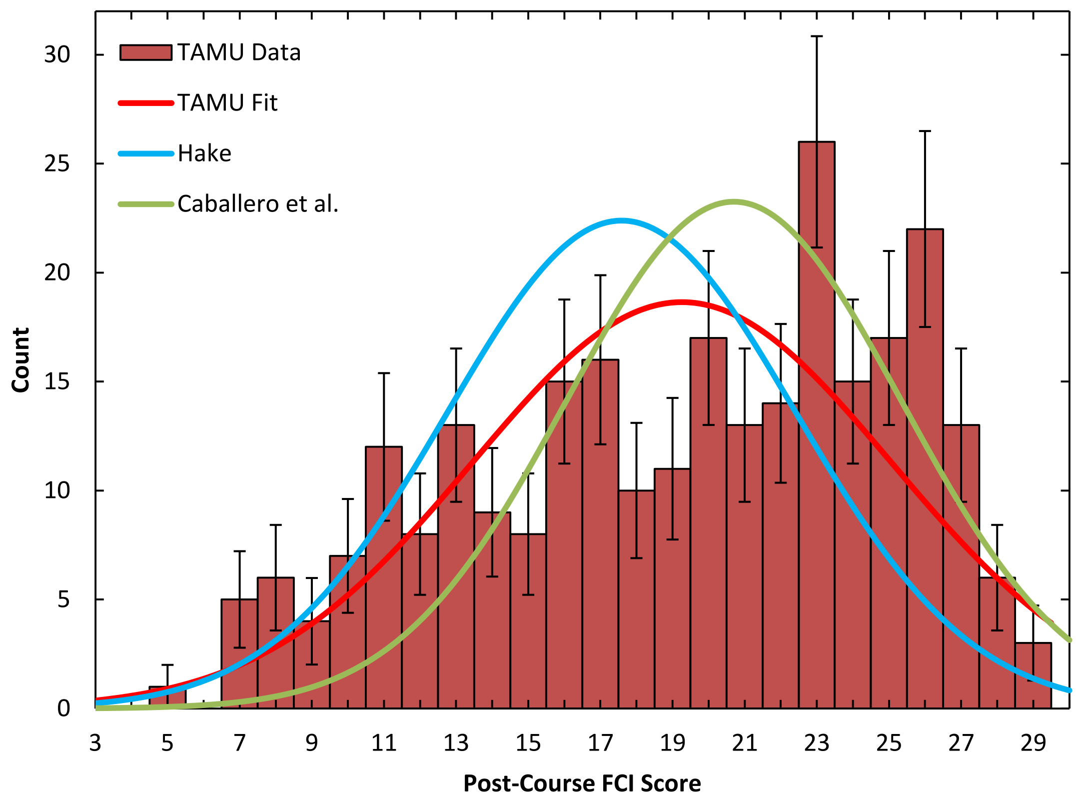

Figure 8. *Post-Course FCI Distribution*. A comparison between data obtained for this study and two other post-course distributions (Hake, Caballero et al.). The non-TAMU fits are scaled to a sample size equivalent to the TAMU fit. The sample statistics for this distribution are shown in Table 6.

Table 6. *Post-Course FCI Distribution Statistics*. The sample statistics for the post-course FCI for this study and two other distributions (Hake and Caballero et al.).

Data Set	*Sample Size* (n)	*Sample Mean* $(\bar{x})$	*Sample Variance* (s^2)
TAMU	271	19.26 ± 0.69	33.61 ± 5.66
Hake	975	17.59 ± 0.30	22.32 ± 2.07
Caballero et al.	1983	20.70 ± 0.20	21.61 ± 1.34

5.3 Outgoing Student Distributions

Mathematically, the outgoing students have increased calculus knowledge with a smaller variance, suggesting that fewer students are missing vital calculus concepts and skills by the end of the course, seen in Figure 7. Performance on the post-course FCI is also typical for an introductory mechanics course. Figure 8 shows the TAMU data compared to the two other outgoing distributions, as previously described.

5.4 Pre-Course and Post-Course FCI t-Test Analysis

As shown in Figure 9 and Table 7, there is a statistically significant increase in performance on the FCI between the beginning and end of the course because the bounds on the p-value do not enclose values greater than $\alpha = .05$. Together, the reported CLES values and the sample means show that any given student will score approximately three additional questions correct on the post-course FCI approximately 71% of the time.

5.5 Statistical Comparisons with Other Treatments

Compared to the distributions of Hake and Caballero et al., seen in Tables 8 through 11, approximately 65% of TAMU students entered the course able to answer about two more questions correct on the FCI than is typical, assuming the Hake distribution is a typical student population. This is a reasonable assumption because the Hake data is collected from many different participating institutions with a very large sample size. Upon leaving the course, approximately 63% of the TAMU students answered about two more questions on the FCI than is typical.

It should be noted here that the TAMU distribution only includes students that completed the course, so the incoming population is skewed toward higher scores, and the Hake distributions are based on the traditional lecture-style college courses contained in the full Hake data set.

In addition, it is very apparent that additional mathematics requirements, in some way, correlate to performing higher on the post-course FCI. Upon leaving the course, approximately 62% of Caballero et al. students were able to answer one additional question correct on the post-course FCI compared to TAMU students. This is significant considering the TAMU students entered the course with higher FCI performance. This connection suggests that performance on the FCI is not entirely based upon a student's understanding of conceptual Newtonian physics.

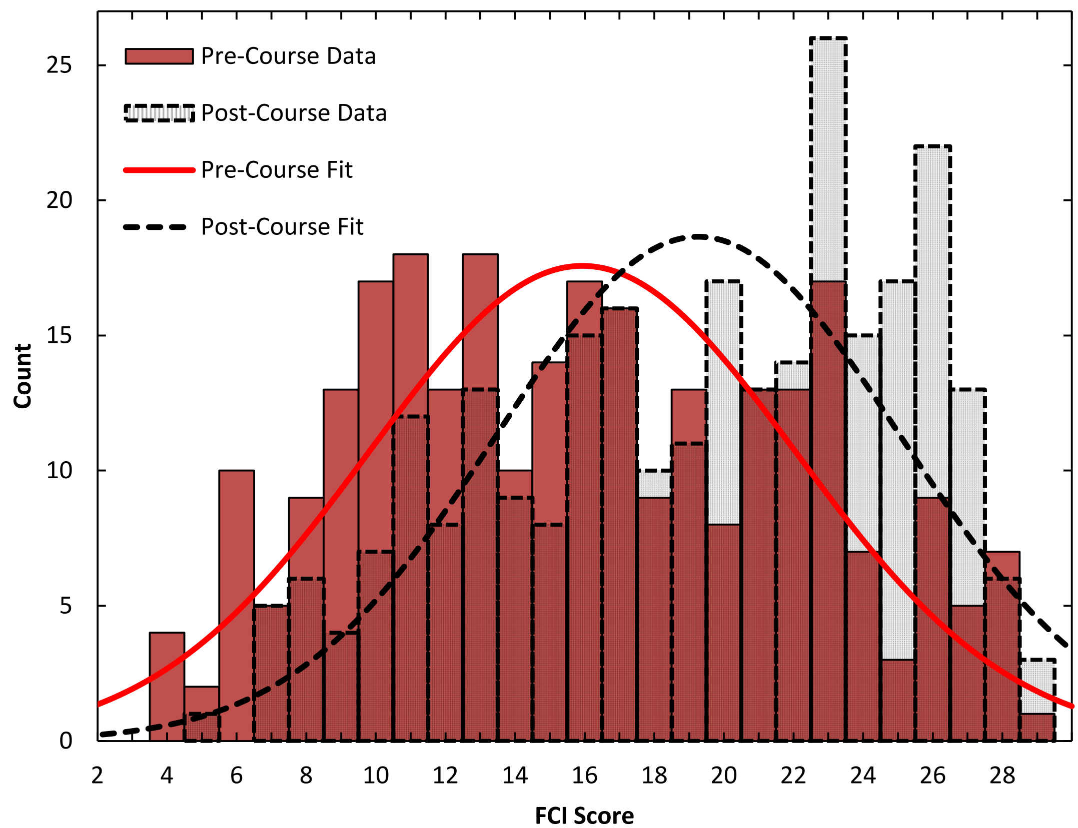

Figure 9. *Pre-Course and Post-Course TAMU FCI Distributions*. Pre-course and post-course FCI distribution t-Test for TAMU data, along with the corresponding Gaussian fits.

Table 7. *Pre-Course and Post-Course TAMU FCI t-Test Results.* The t-Test results comparing the pre-course and post-course FCI performance for the TAMU data set.

	t-value	*dof*	*p-value*	*Effect Size*	*CLES*
Upper Bound	10.15		$< 10^{-8}$ %	0.872	80.8%
Central Value	6.48	538	2.05×10^{-8} %	0.557	71.1%
Lower Bound	3.43		6.41×10^{-2} %	0.295	61.6%

Table 8. *Caballero et al. and TAMU Pre-Course FCI t-Test Results.* The t-Test results comparing the pre-course FCI performance for the Caballero et al. and TAMU data sets. The two Gaussian fits can be seen in Figure 6.

	t-value	*dof*	*p-value*	*Effect Size*	*CLES*
Upper Bound	8.20	304	$< 10^{-8}$ %	0.610	72.9%
Central Value	5.15	300	4.66×10^{-5} %	0.404	65.7%
Lower Bound	2.59	297	0.994 %	0.212	58.4%

Table 9. *Hake and TAMU Pre-Course FCI t-Test Results.* The t-Test results comparing the pre-course FCI performance for the Hake and TAMU data sets. The two Gaussian fits can be seen in Figure 6.

	t-value	*dof*	*p-value*	*Effect Size*	*CLES*
Upper Bound	7.86	374	$< 10^{-8}$ %	0.606	72.8%
Central Value	4.66	365	4.37×10^{-4} %	0.369	64.4%
Lower Bound	1.96	358	5.09%	0.158	56.3%

Table 10. *TAMU and Caballero et al. Post-Course FCI t-Test Results.* The t-Test results comparing the post-course FCI performance for the TAMU and Caballero et al. data sets. The two Gaussian fits can be seen in Figure 8.

	t-value	*dof*	*p-value*	*Effect Size*	*CLES*
Upper Bound	6.92	325	$< 10^{-8}$ %	0.506	69.4%
Central Value	3.91	319	1.13×10^{-2} %	0.299	61.8%
Lower Bound	1.37	314	17.2%	0.108	54.3%

Table 11. *Hake and TAMU Post-Course FCI t-Test Results.* The t-Test results comparing the post-course FCI performance for the Hake and TAMU data sets. The two Gaussian fits can be seen in Figure 8.

	t-value	*dof*	*p-value*	*Effect Size*	*CLES*
Upper Bound	7.55	391	$< 10^{-8}$ %	0.560	71.2%
Central Value	4.36	380	1.68×10^{-3} %	0.332	63.0%
Lower Bound	1.65	372	9.92%	0.128	55.1%

5.6 FCI-Midterm Correlation

The average midterm score distribution is relatively typical for an introductory mechanics course, shown in Figure 10. An analysis of the pre-course FCI and midterm exam correlation should highlight the initial state of the student's inherent conceptual reasoning abilities. There is approximately a 14% correlation between overall student performance in the course and their performance on the FCI, seen in Figure 11 and Table 13. In particular, there is approximately an 18% correlation between the first midterm exam and the FCI, shown in Figure 12 and Table 14. This is expected because the FCI and the first midterm both cover mostly kinematics and early applications of Newton's laws. There is only about a 3% correlation between the FCI and the second midterm, seen in Figure 13 and Table 15, and about an 8% correlation between the FCI and the third midterm, seen in Figure 14 and Table 16. The correlations here are mostly expected, since these exams do not share much content with the FCI. However, these two correlations are expected to be about the same and are not. The reasoning behind this is not entirely clear.

All of the same comparisons can be made with the post-course FCI with very similar correlations, shown in Figures 15-18 and Tables 17-20. This suggests that, regardless of *when* the student takes the FCI, the correlation remains the same. It is expected that, if a student learns the concepts presented in an introductory physics course, there should be a higher explained variance between course performance and FCI performance. Statistically speaking, shown in Figure 19 and Table 21, correlation does *not* change at all between the pre-course and post-course FCI. This suggests that increased performance on the FCI is *not* due to an increased conceptual understanding of Newtonian mechanics and is, instead, due to some other effect, such as educational background, student intelligence, or increased test-taking abilities. Further study would be needed to come to more exact conclusion.

5.7 Incoming Mathematics Correlation

There is also a connection between a student's incoming mathematics skills and their performance in the course, shown in Figures 20-21 and Tables 22-23. Not only is there an approximate 11% correlation between a student's pre-course MIE score and their overall course performance, but about 75% of incoming students that perform well on the MIE score about 10% higher in the course, compared to their poorly performing counterparts. This suggests that mathematics pre-requisites are of great importance for introductory mechanics courses and should be *increased in difficulty*, such as requiring two semester of calculus like the courses involved in the Caballero et al. study. In addition, students entering the course with higher MIE scores perform better on the FCI and perform similarly to students from the Caballero study, despite the lack of mathematics on the FCI, shown in Figures 22-23 and Tables 24-25.

These results suggest that students with a more complete mathematics education are either more educated in general, possess a greater intelligence, or simply have better reasoning and test-taking skills. Regardless of exact causation, requiring students to have completed additional mathematics coursework before they enroll in their first introductory physics course will increase student performance within the course.

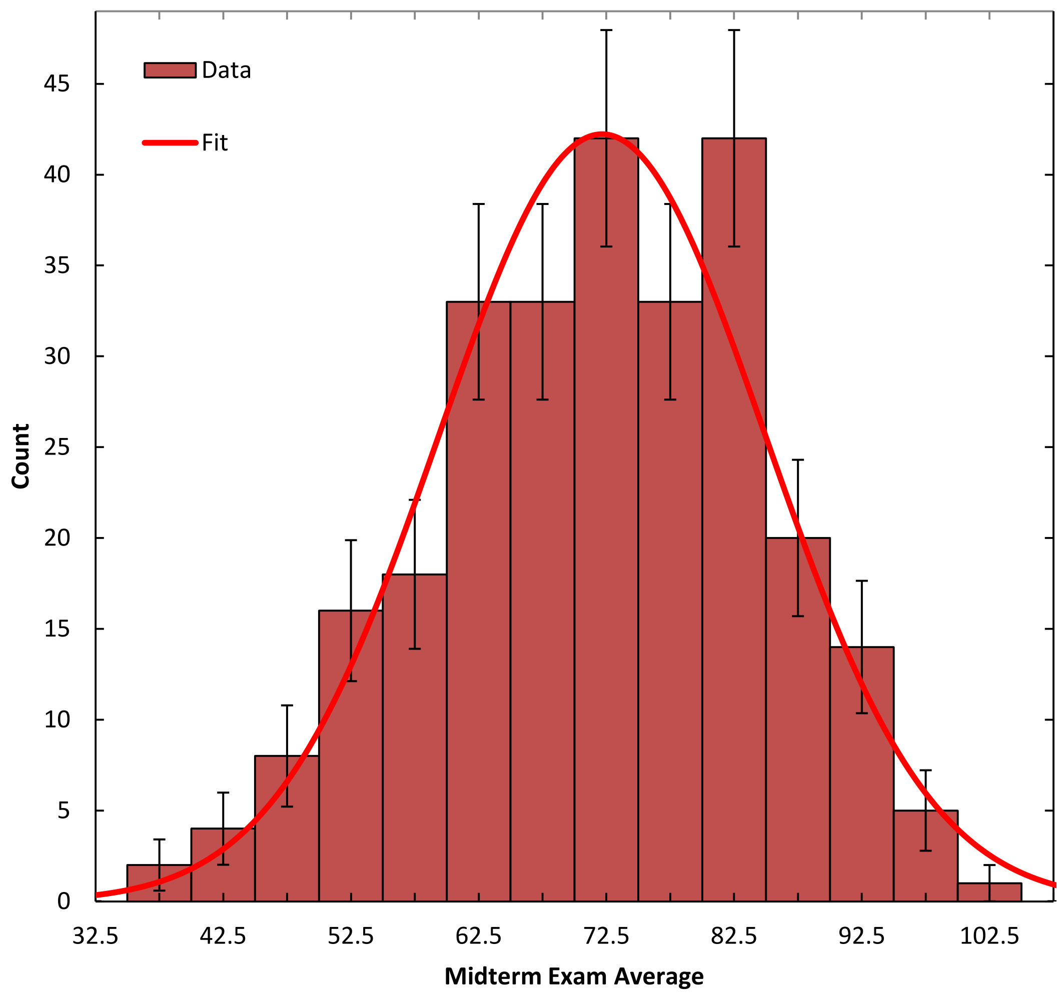

Figure 10. *Midterm Exam Average Distribution*. The midterm exam average distribution and corresponding Gaussian fit for the TAMU data. The sample statistics for this distribution are shown in Table 12.

Table 12. *Midterm Average Distribution Statistics*. The sample statistics for the average on all three midterm examinations for this study.

Sample Size (n)	*Sample Mean* $(\bar{x})$	*Sample Variance* (s^2)
271	72.16 ± 1.52	163.86 ± 27.59

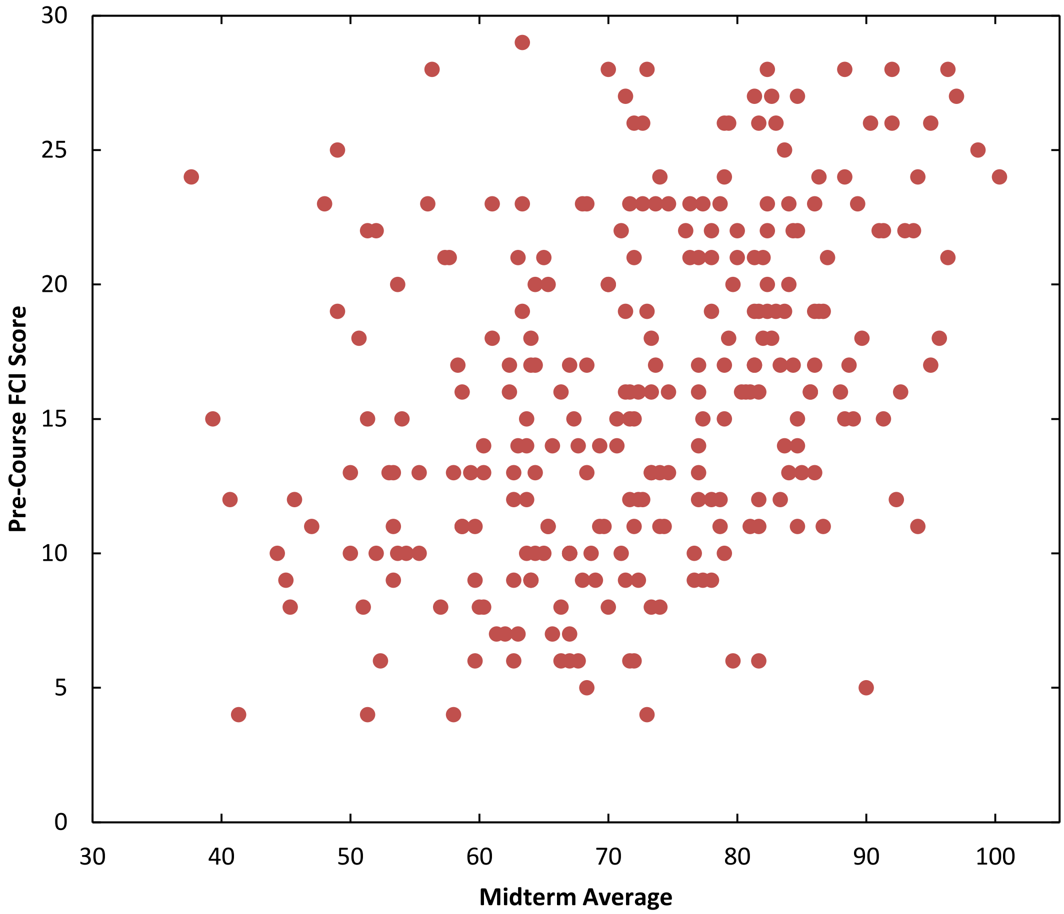

Figure 11. *Pre-Course FCI and Midterm Average Correlation.* The correlation plot showing each student's pre-course FCI score and average midterm score.

Table 13. *Pre-Course FCI and Midterm Average Regression Analysis.* The regression analysis for the pre-course FCI and average midterm exam correlation.

dof = 268	*r*	*t-value*	*p-value*	r^2
Upper Bound	0.480	8.97	$< 10^{-8}$ %	23.1%
Central Value	0.369	6.51	3.72×10^{-8} %	13.6%
Lower Bound	0.258	4.38	1.72×10^{-3} %	6.68%

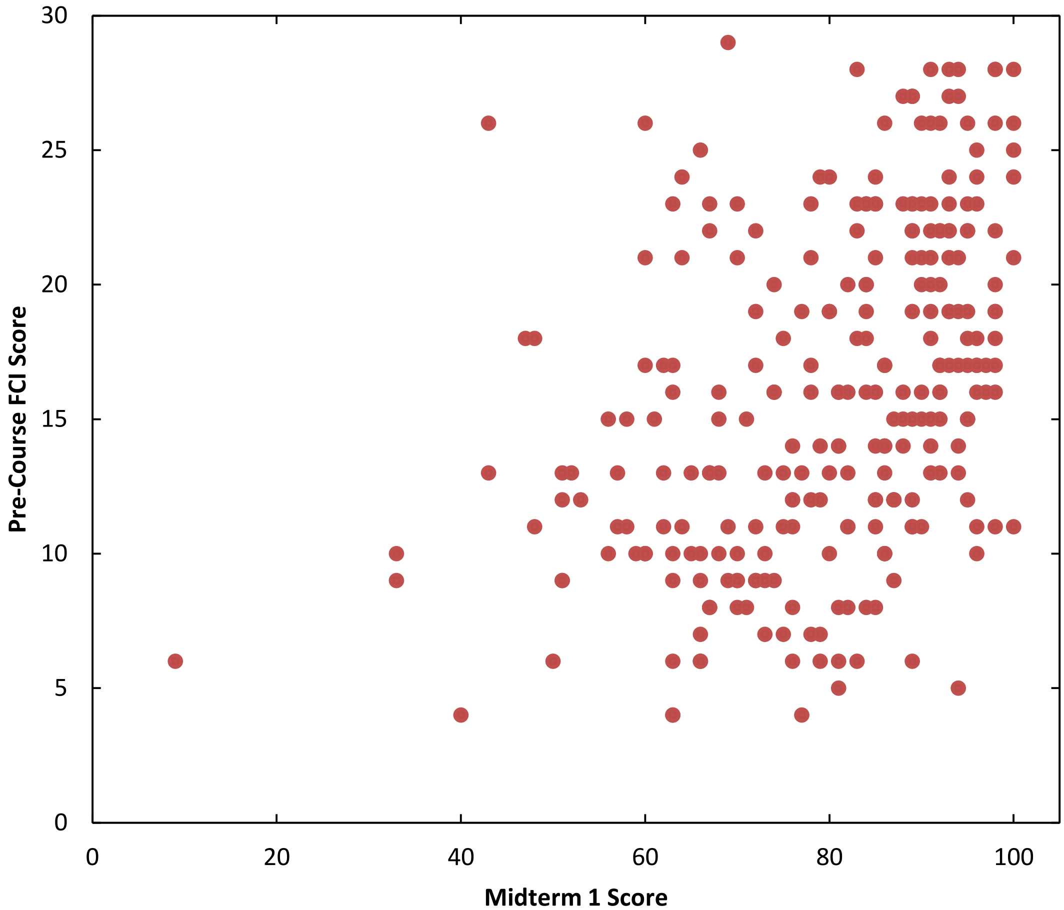

Figure 12. *Pre-Course FCI and Midterm 1 Correlation*. The correlation plot showing each student's pre-course FCI score and first midterm score.

Table 14. *Pre-Course FCI and Midterm 1 Regression Analysis*. The regression analysis for the pre-course FCI and first midterm correlation.

dof = 268	*r*	*t-value*	*p-value*	r^2
Upper Bound	0.528	10.2	$< 10^{-8}$ %	27.9%
Central Value	0.419	7.56	$< 10^{-8}$ %	17.6%
Lower Bound	0.311	5.36	1.84×10^{-5} %	9.67%

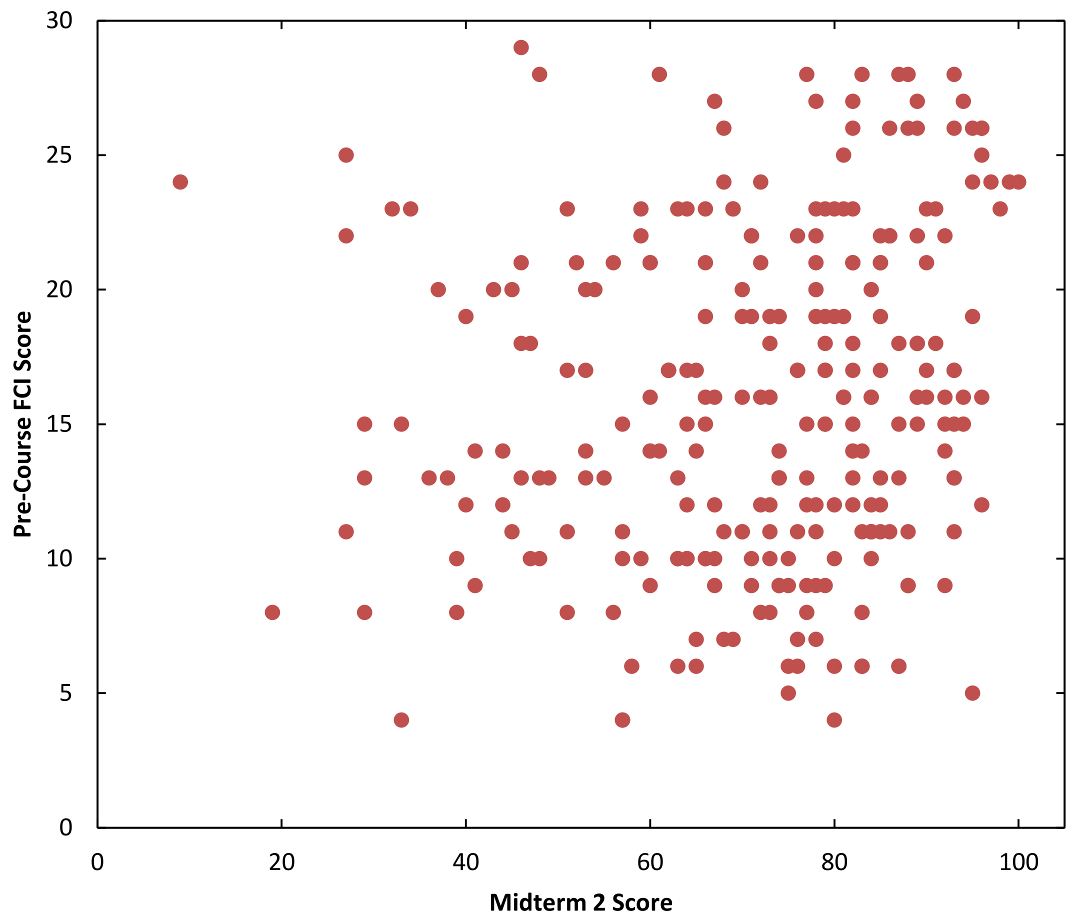

Figure 13. *Pre-Course FCI and Midterm 2 Correlation*. The correlation plot showing each student's pre-course FCI score and second midterm score.

Table 15. *Pre-Course FCI and Midterm 2 Regression Analysis.* The regression analysis for the pre-course FCI and second midterm correlation.

dof = 268	*r*	*t-value*	*p-value*	r^2
Upper Bound	0.268	4.56	7.90×10^{-4} %	7.19%
Central Value	0.150	2.48	1.36%	2.25%
Lower Bound	0.0318	0.522	60.2%	.101%

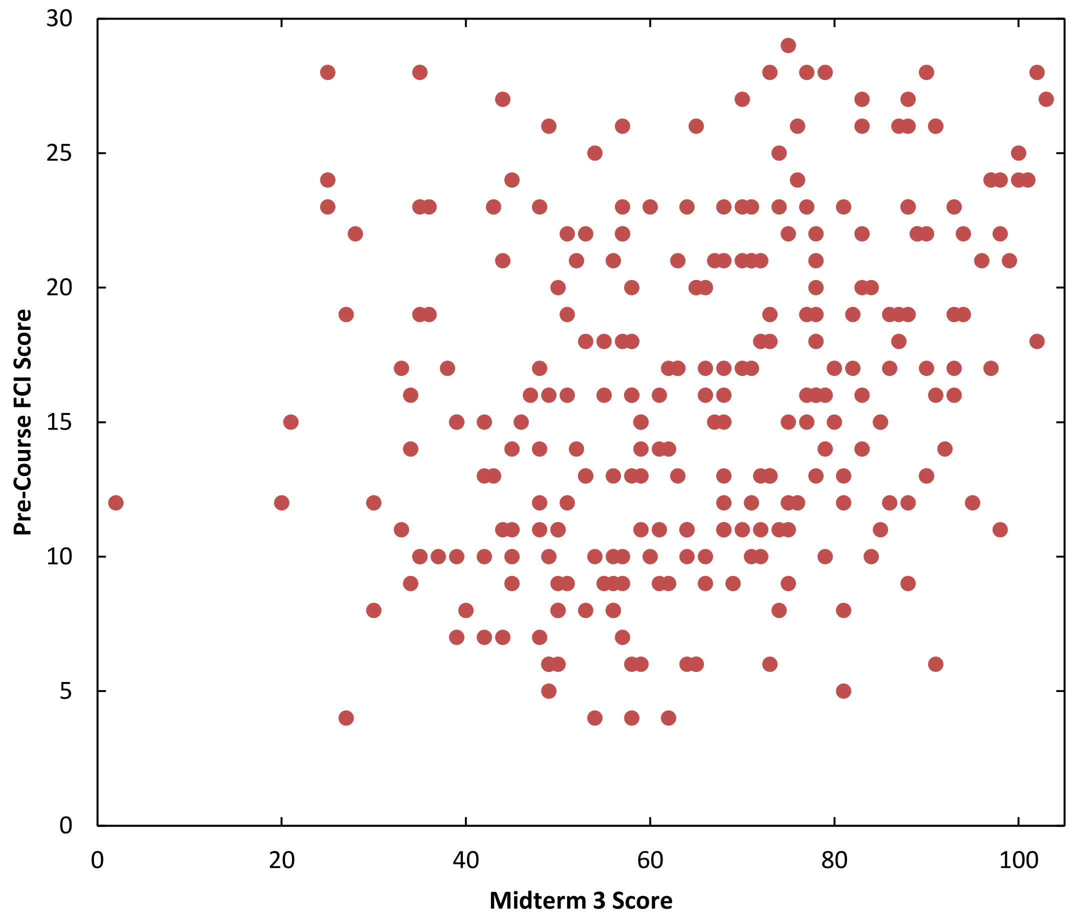

Figure 14. *Pre-Course FCI and Midterm 3 Correlation*. The correlation plot showing each student's pre-course FCI score and third midterm score.

Table 16. *Pre-Course FCI and Midterm 3 Regression Analysis*. The regression analysis for the pre-course FCI and third midterm correlation.

dof = 268	*r*	*t-value*	*p-value*	r^2
Upper Bound	0.394	7.02	$< 10^{-8}$ %	15.5%
Central Value	0.279	4.76	3.10×10^{-4} %	7.81%
Lower Bound	0.165	2.73	0.667%	2.71%

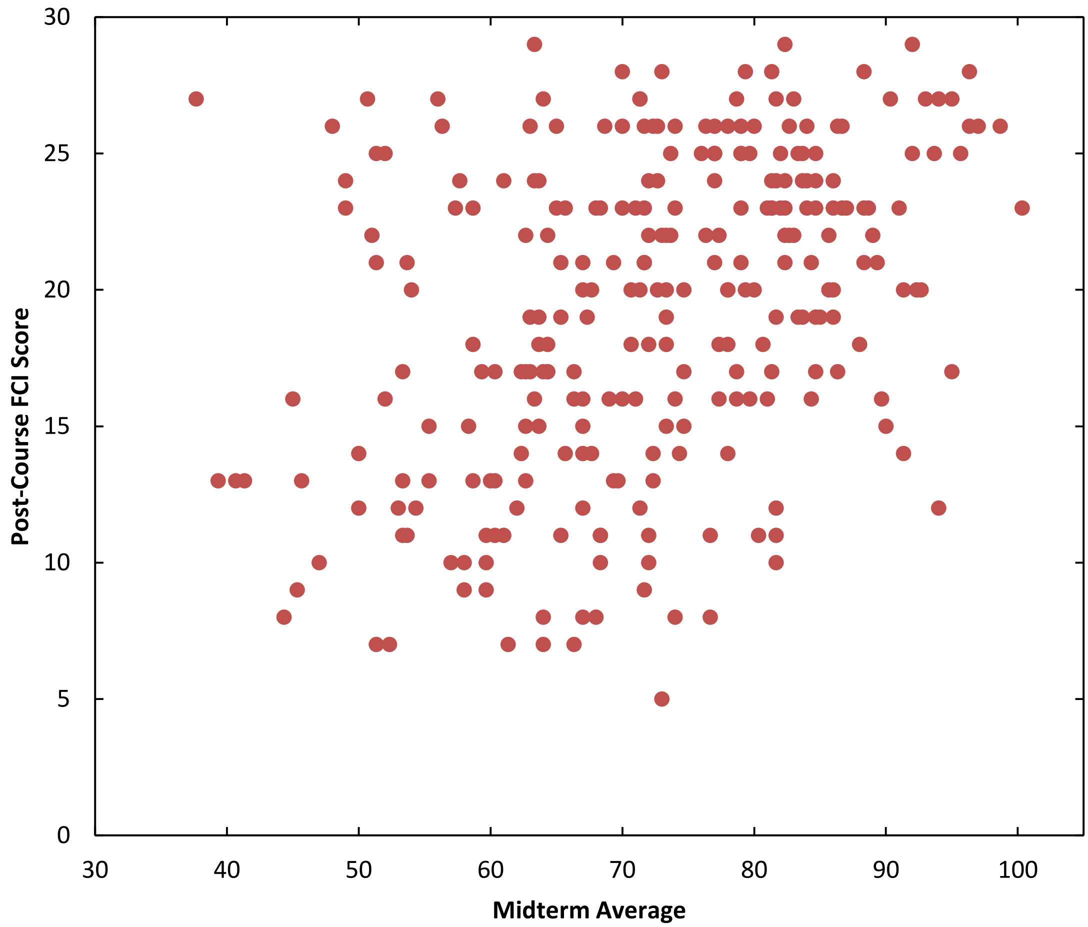

Figure 15. *Post-Course FCI and Midterm Average Correlation*. The correlation plot showing each student's post-course FCI score and average midterm score.

Table 17. *Post-Course FCI and Midterm Average Regression Analysis.* The regression analysis for the post-course FCI and average midterm exam correlation.

dof = 268	*r*	*t-value*	*p-value*	r^2
Upper Bound	0.500	9.44	$< 10^{-8}$ %	25.0%
Central Value	0.389	6.92	$< 10^{-8}$ %	15.2%
Lower Bound	0.279	4.76	3.11×10^{-4} %	7.81%

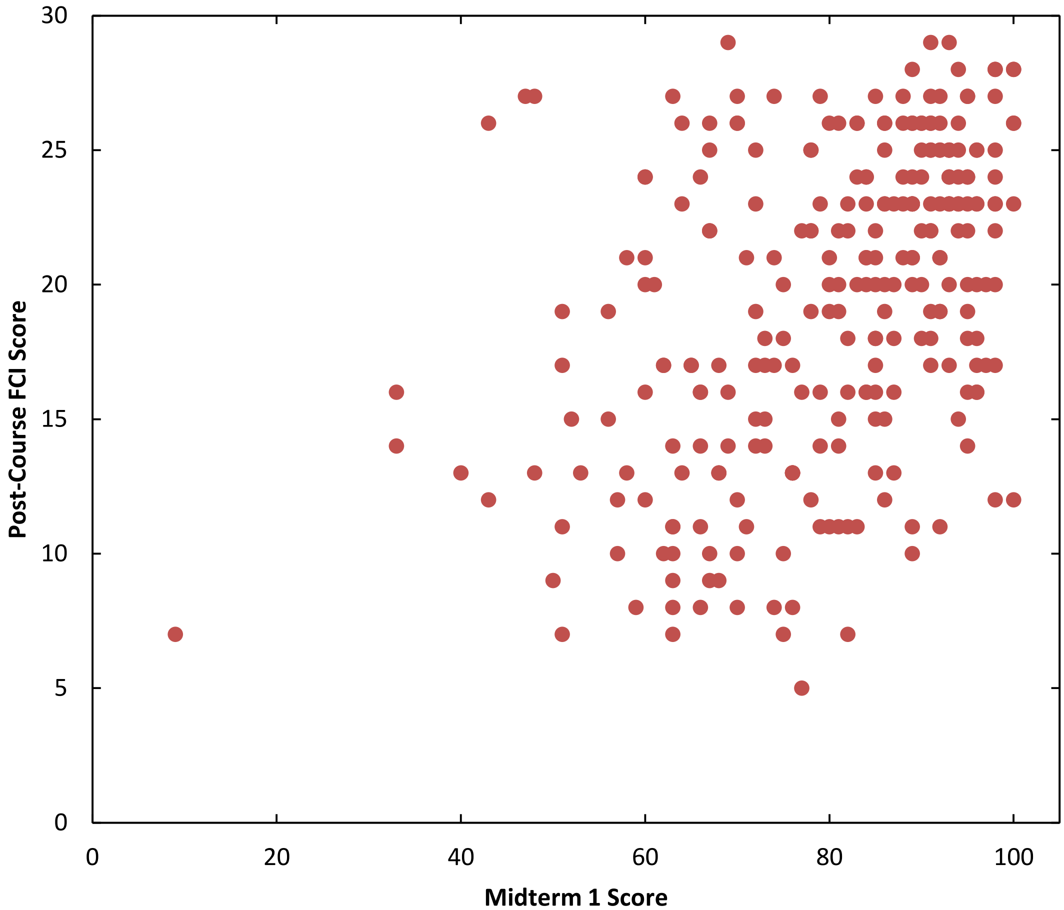

Figure 16. *Post-Course FCI and Midterm 1 Correlation*. The correlation plot showing each student's post-course FCI score and first midterm score.

Table 18. *Post-Course FCI and Midterm 1 Regression Analysis*. The regression analysis for the post-course FCI and first midterm correlation.

dof = 268	*r*	*t-value*	*p-value*	r^2
Upper Bound	0.541	10.5	$< 10^{-8}$ %	29.3%
Central Value	0.434	7.88	$< 10^{-8}$ %	18.8%
Lower Bound	0.326	5.65	4.20×10^{-6} %	10.6%

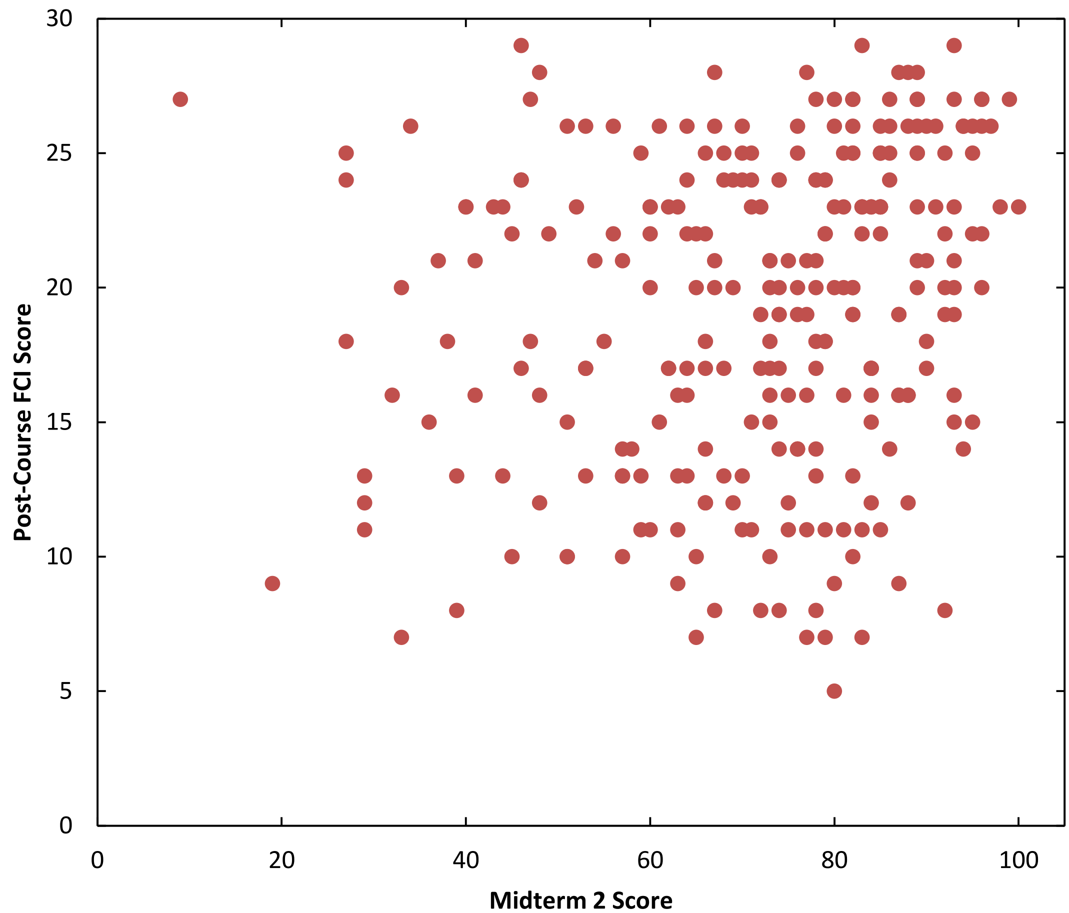

Figure 17. *Post-Course FCI and Midterm 2 Correlation*. The correlation plot showing each student's post-course FCI score and second midterm score.

Table 19. *Post-Course FCI and Midterm 2 Regression Analysis.* The regression analysis for the post-course FCI and second midterm correlation.

dof = 268	*r*	*t-value*	*p-value*	r^2
Upper Bound	0.308	5.31	2.33×10^{-5} %	9.51%
Central Value	0.191	3.19	0.160%	3.65%
Lower Bound	0.0738	1.21	22.7%	0.545%

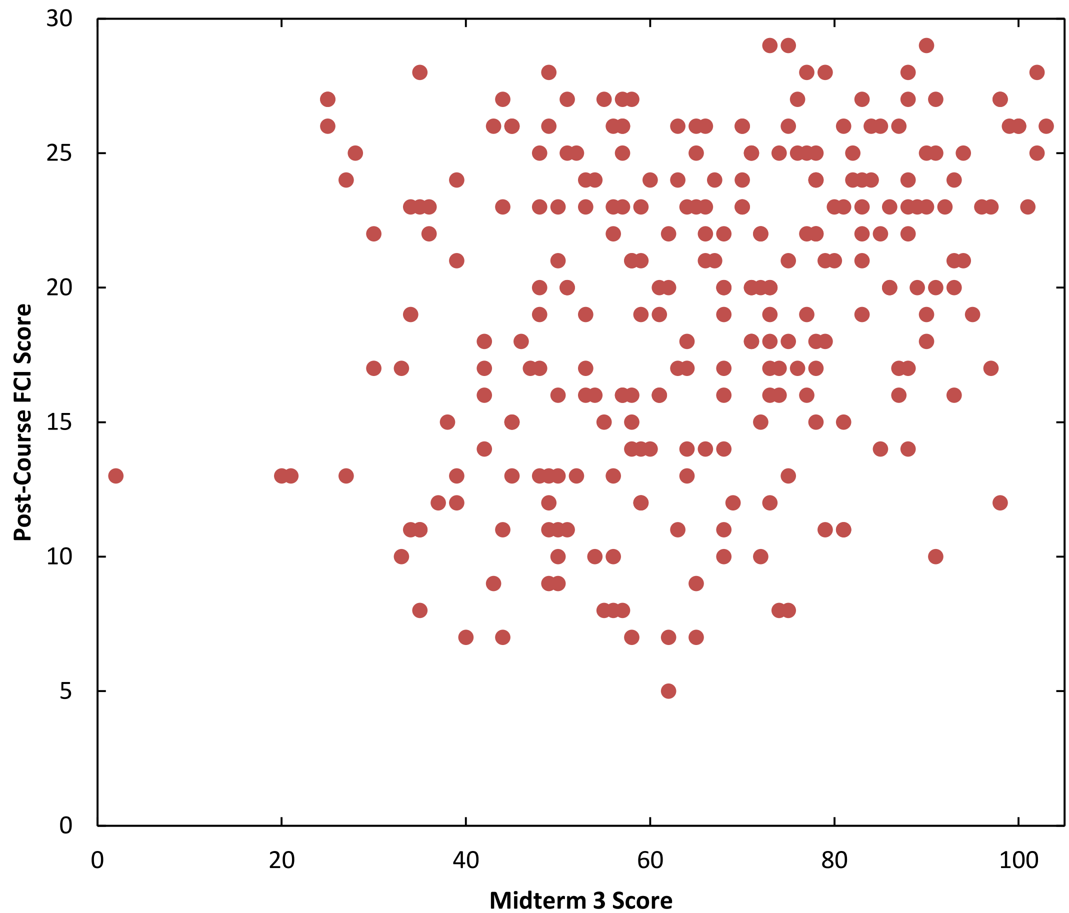

Figure 18. *Post-Course FCI and Midterm 3 Correlation*. The correlation plot showing each student's post-course FCI score and third midterm score.

Table 20. *Post-Course FCI and Midterm 3 Regression Analysis*. The regression analysis for the post-course FCI and third exam correlation.

dof = 268	*r*	*t-value*	*p-value*	r^2
Upper Bound	0.386	6.85	$< 10^{-8}$ %	15.5%
Central Value	0.271	4.61	3.10×10^{-4} %	7.81%
Lower Bound	0.156	2.58	0.667%	2.71%

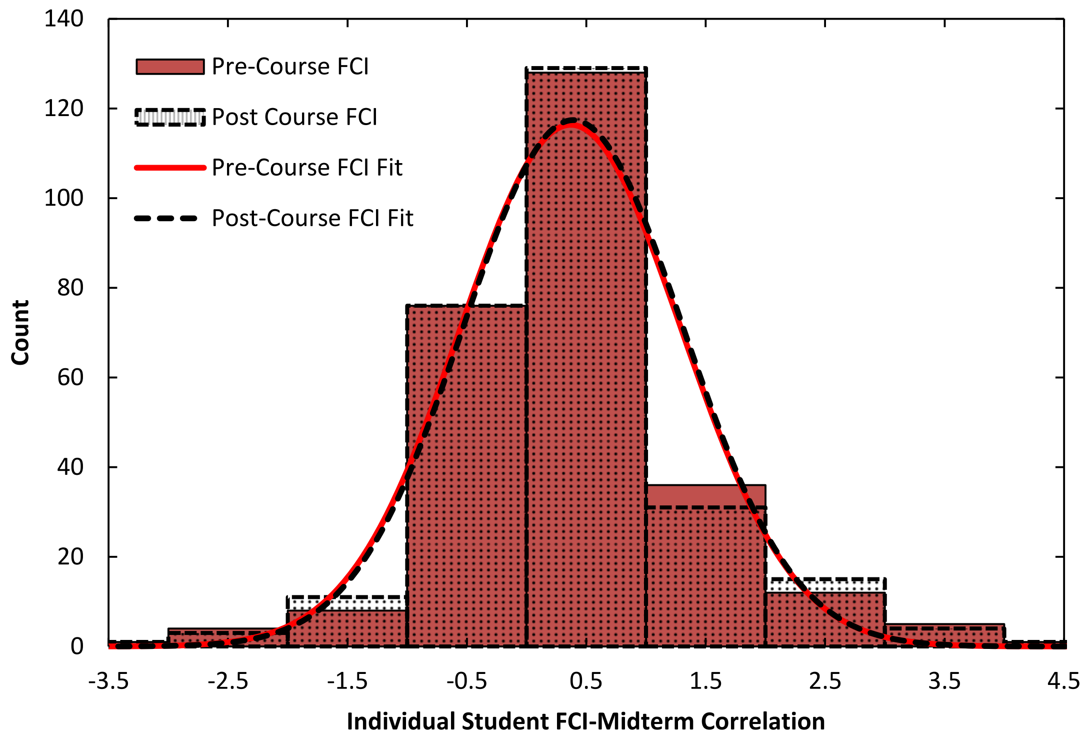

Figure 19. *Individual Student FCI-Midterm Correlation Distributions*. The distribution of each individual student's pre- and post- FCI-Midterm correlation.

Table 21. *Individual Student FCI-Midterm Correlation Statistics and t-Test Results*. The sample statistics and t-Test results comparing pre- and post- FCI-Midterm correlation.

Data Set	*Sample Size* (n)	*Sample Mean* $(\bar{r})$	*Sample Variance* (s^2)
Pre-Course	271	0.369 ± 0.111	0.864 ± 0.146
Post-Course	271	0.389 ± 0.110	0.848 ± 0.144

	t-value	*dof*	*p-value*	*Effect Size*	*CLES*
Upper Bound	3.32		0.0974%	0.287	61.3%
Central Value	0.251	533	80.2%	0.0217	50.9%
Lower Bound	-2.33		2.01%	-0.201	42.0%

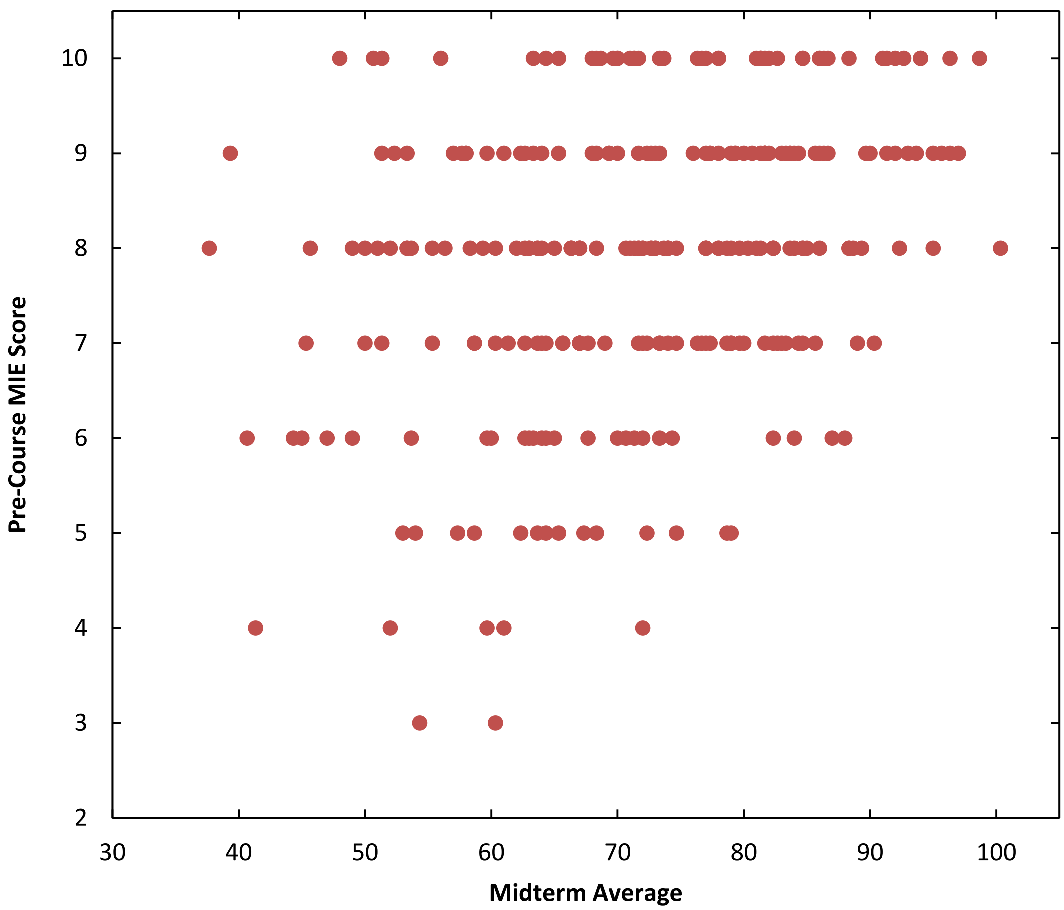

Figure 20. *Pre-Course MIE and Midterm Average Correlation*. The correlation plot showing each student's pre-course MIE score and average midterm score.

Table 22. *Pre-Course MIE and Midterm Average Regression Analysis.* The regression analysis for the pre-course MIE and average exam correlation.

dof = 268	*r*	*t-value*	*p-value*	r^2
Upper Bound	0.451	8.27	$< 10^{-8}$ %	20.3%
Central Value	0.338	5.89	1.17×10^{-6} %	11.5%
Lower Bound	0.226	3.80	1.81×10^{-2} %	5.11%

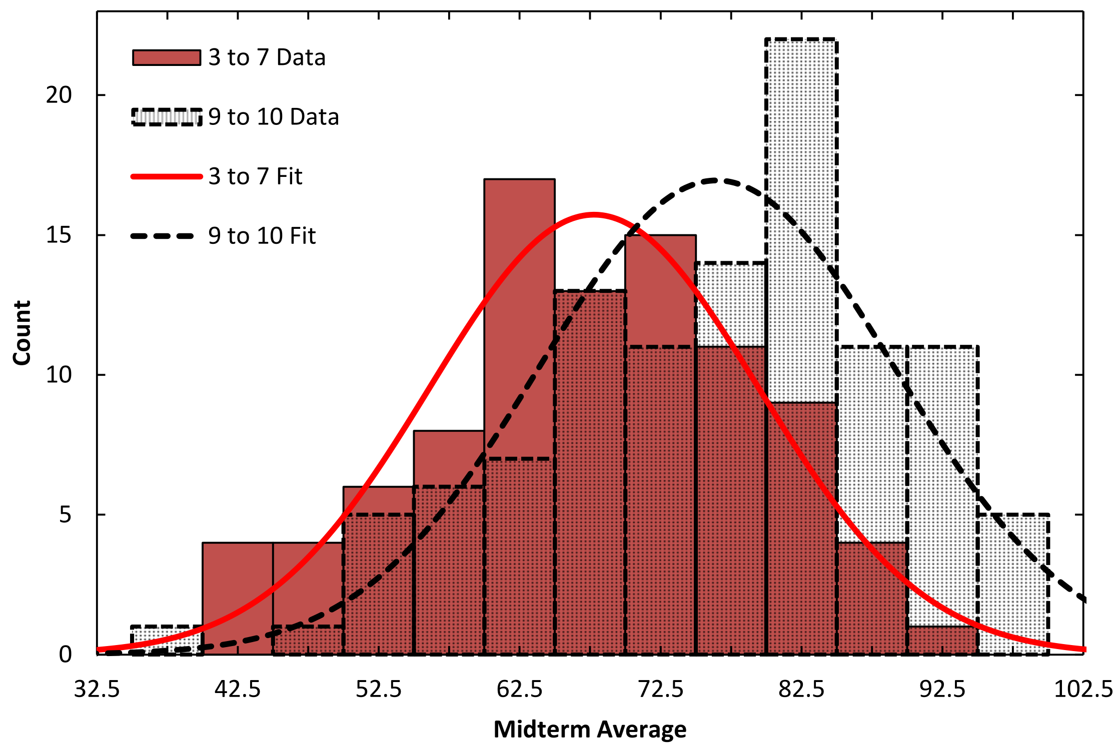

Figure 21. *Midterm Exam Average Distributions by Pre-Course MIE*. The midterm average distributions for two student groups, separated by MIE performance.

Table 23. *Midterm Exam Average by Pre-Course MIE Statistics and t-Test Results*. The distribution statistics and t-Test results for the two student groups.

Data Set	*Sample Size* (n)	*Sample Mean* ($\bar{x}$)	*Sample Variance* (s^2)
MIE 3 to 7	92	67.78 ± 2.39	136.25 ± 39.37
MIE 9 to 10	107	76.48 ± 2.39	158.51 ± 42.47

	t-value	*dof*	*p-value*	*Effect Size*	*CLES*
Upper Bound	9.22	196	$< 10^{-8}$ %	1.31	90.3%
Central Value	5.06	195	9.79×10^{-5} %	0.715	76.3%
Lower Bound	2.02	195	4.46%	0.286	61.3%

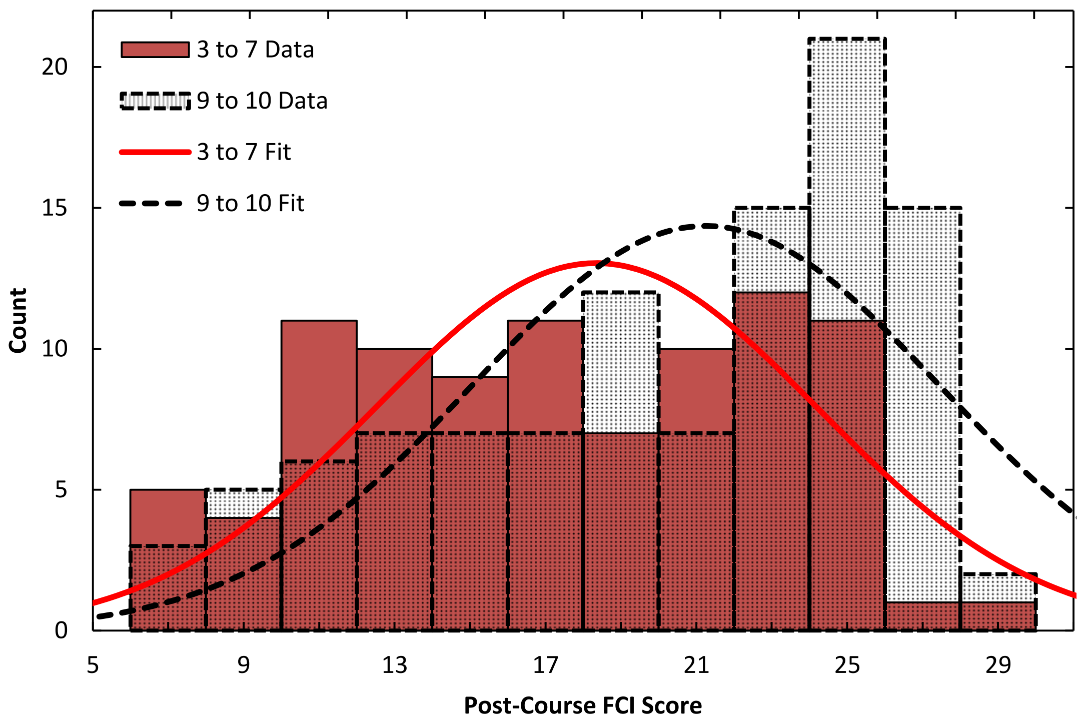

Figure 22. *Post-Course FCI Distributions by Pre-Course MIE.* The post-course FCI distributions for two student groups, separated by MIE performance.

Table 24. *Post-Course FCI by Pre-Course MIE Statistics and t-Test Results.* The distribution statistics and t-Test results for the two student groups.

Data Set	***Sample Size (n)***	***Sample Mean ($\bar{x}$)***	***Sample Variance (s^2)***
MIE 3 to 7	92	17.84 ± 1.15	31.68 ± 9.15
MIE 9 to 10	107	20.63 ± 1.13	35.37 ± 9.48

	t-value	***dof***	***p-value***	***Effect Size***	***CLES***
Upper Bound	7.26	195	$< 10^{-8}$ %	1.03	84.8%
Central Value	3.40	195	8.31×10^{-2} %	0.481	68.5%
Lower Bound	0.551	194	58.2 %	7.81×10^{-2}	53.1%

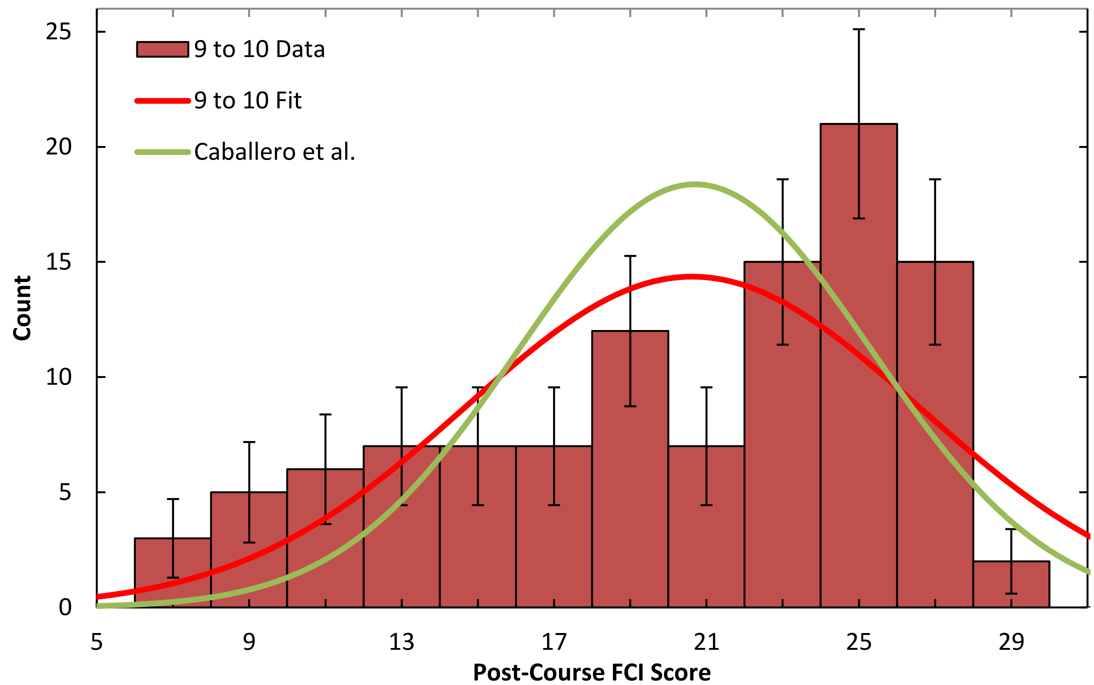

Figure 23. *TAMU MIE 9 to 10 Post-Course FCI Distribution with Caballero et al.*. The post-course FCI distribution for well-performing MIE TAMU students, along with the corresponding Gaussian fit and Caballero et al. fit. The Caballero et al. fit is scaled to the same sample size as the TAMU sample size.

Table 25. *TAMU MIE 9 to 10 Post-Course FCI with Caballero et al. t-Test Results.* The t-Test results comparing the post-course FCI performance of well-performing MIE TAMU students and Caballero et al. students.

	t-value	*dof*	*p-value*	*Effect Size*	*CLES*
Upper Bound	2.79	115	0.613%	0.0152	62.2%
Central Value	0.123	113	90.3%	0.310	50.6%
Lower Bound	-1.92	111	5.74 %	-0.257	39.9%

6. SUMMARY AND CONCLUSIONS

Physics education research is paramount to, not only the success of students, but to the continued success of all physics departments. In recent years, the FCI has become the primary, and sometimes *only*, tool used to evaluate teaching methods and analyze overall student performance. It is unclear whether the FCI is an appropriate metric when used to compare a variety of introductory mechanics courses. It is well understood that mathematics plays a crucial function in a physics student's success or failure, but the exact details of that function are not entirely clear in the context of introductory mechanics.

Traditionally, analysis of teaching methods involves comparing the average gains on the FCI for two treatment groups, each consisting of many classroom-sized samples. This work is unique because it applies a complete statistical treatment to a single sample of large sample size in order to analyze the relationships for each individual student.

It has been shown that the FCI does *not* adequately measure student comprehension within the context of PHYS 218 STEP at TAMU, despite statistically significant gains on the FCI. In addition, a student's mathematics background is of critical importance. It will often mean the difference of an entire letter grade for the average student. This suggests that a different metric for determining the success of an introductory mechanics course should be devised and that mathematics pre-requisites for introductory mechanics courses should be increased.

REFERENCES

1 D. Hestenes, M. Wells, and G. Swackhamer, "Force Concept Inventory," *The Physics Teacher*. **30**, 141-158 (1992).

2 R. Beichner, "An Introduction to Physics Education Research," *Getting Started in PER* (2009), Reviews in PER Vol. 2. Retrieved April 16, 2013, from http://www.per-central.org/items/detail.cfm?ID=8806.

3 D. Hestenes and M. Wells, "A Mechanics Baseline Test," *The Physics Teacher*. **30**, 159 (1992).

4 R. Beichner, R. Hake, L. McDermott, J. Mestre, E. Redish, F. Reif, and J. Risley, "Support of Physics-Education Research as a Subfield of Physics: Proposal to the NSF Physics Division," (1994).

5 R. Hake, "Interactive Engagement Versus Traditional Methods: A Six Thousand Student Survey of Mechanics Test Data for Introductory Physics Courses," *American Journal of Physics*. **66** (1), 64-78 (1998).

6 D. Huffman and P. Heller, "What Does the Force Concept Inventory Actually Measure?" *The Physics Teacher*. **33**, 138-143 (1995).

7 D. Hestenes and I. Halloun, "Interpreting the Force Concept Inventory: A Response to March 1995 Critique by Huffman and Heller," *The Physics Teacher*. **33**, 502-506 (1995).

8 P. Heller and D. Hestenes, "Interpreting the Force Concept Inventory: A Reply to Hestenes and Halloun," *The Physics Teacher*. **33**, 503-511 (1995).

9 W. Bassichis, *Don't Panic – Volume 1, Mechanics* (4th ed.). College Station, TX: OR Publishing (2007).

10 Student, "The Probable Error of a Mean," *Biometrika*. **6** (1), 1-25 (1908).

11 B.L. Welch, "The Generalization of Student's Problem when Several Different Population Variances are Involved," *Biometrika*. **34** (1), 28-35 (1947).

12 F.E. Satterthwaite, "An Approximate Distribution of Estimates of Variance Components," *Biometrics Bulletin*. **2** (6), 110-114 (1946).

13 J. Cohen, "Statistical Power Analysis," *Current Directions in Psychological Science*. **1** (3), 98-101 (1992).

14 J. Ruscio, "A Probability-Based Measure of Effect Size: Robustness to Base Rates and Other Factors," *Psychological Methods*. **13** (1), 19-30 (2008).

15 K. Pearson, "Mathematical Contributions to the Theory of Evolution. III. Regression, Heredity, and Panmixia," *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*. **187**, 253-318 (1896).

16 M. Caballero et al., E. Murray, K. Bujak, M. Marr et al., "Comparing Large Lecture Mechanics Curricula using the Force Concept Inventory: A Five Thousand Student Study," *The American Journal of Physics*. **80** (7), 638-644 (2012).

APPENDIX

ID	Section	ME 1	ME 2	ME 3	Pre MIE	Pre MIE (Non-Calc)	Pre MIE (Calc)	Post MIE (Calc)	Pre FCI	Post FCI
1	10	87	88	57	9	5	4	4	9	16
2	1	91	60	44	6	4	2	5	21	23
3	12	89	85	70	8	4	4	4	21	23
4	18	95	96	86	8	4	4	5	12	20
5	19	58	57	39	10	5	5	4	15	21
6	3	33	57	60	8	5	3	3	10	14
7	6	98	84	65	6	5	1	4	20	23
8	10	80	87	53	6	2	4	4	13	19
9	10	91	92	48	8	5	3	4	14	25
10	20	63	66	58	9	4	5	5	16	14
11	19	82	60	34	5	3	2	4	16	23
12	4	75	55	90	9	5	4	4	13	18
13	18	85	99	98	10	5	5	5	24	27
14	16	90	73	79	9	5	4	4	16	18
15	13	72	40	35	8	5	3	5	19	23
16	5	92	92	61	9	5	4	5	16	19
17	3	91	70	78	8	5	3	4	20	25
18	19	88	81	36	10	5	5	4	23	23
19	19	76	72	74	8	3	5	4	8	8
20	5	90	85	85	10	5	5	4	11	26
21	17	84	73	58	8	4	4	4	16	21
22	14	94	44	59	7	4	3	2	14	23
23	19	63	80	25	10	5	5	3	23	27
24	1	57	51	33	6	4	2	1	11	10
25	21	85	78	71	8	5	3	5	12	18
26	5	81	74	34	8	4	4	4	14	19
27	2	62	46	73	3	2	1	3	13	17
28	18	78	66	72	7	4	3	5	21	22
29	2	67	19	50	7	5	2	2	8	9
30	6	96	63	54	10	5	5	4	10	16
31	7	84	73	53	6	4	2	5	8	16
32	17	50	80	49	9	5	4	4	6	9
33	10	93	93	90	10	5	5	5	28	29
34	1	72	76	48	9	5	4	5	11	19
35	4	60	68	88	8	4	4	5	26	24

ID	Section	ME 1	ME 2	ME 3	Pre MIE	Pre MIE (Non-Calc)	Pre MIE (Calc)	Post MIE (Calc)	Pre FCI	Post FCI
36	11	66	76	59	7	4	3	3	6	14
37	3	94	94	103	9	5	4	4	27	26
38	12	59	39	35	6	4	2	5	10	8
39	4	91	66	87	10	5	5	3	19	17
40	3	90	95	91	9	5	4	4	26	25
41	1	93	68	76	7	5	2	4	24	25
42	21	73	78	66	5	3	2	4	9	14
43	18	77	77	68	8	5	3	3	13	16
44	2	70	75	69	6	4	2	4	9	12
45	16	92	62	90	9	5	4	4	17	23
46	13	96	90	78	6	4	2	5	16	18
47	13	94	85	52	8	4	4	4	21	25
48	4	77	95	77	7	4	3	4	19	22
49	18	92	85	51	9	5	4	4	22	25
50	10	66	78	57	7	5	2	4	7	8
51	12	82	45	66	5	4	1	4	20	22
52	9	93	72	96	6	5	1	4	21	23
53	19	81	75	49	5	4	1	3	5	11
54	13	98	85	36	8	5	3	4	19	22
55	6	100	89	49	9	4	5	4	26	28
56	10	63	87	65	8	5	3	4	6	9
57	20	95	89	89	10	5	5	5	22	23
58	9	87	82	48	9	5	4	5	12	13
59	7	100	90	99	9	5	4	5	21	26
60	3	51	74	53	8	5	3	3	13	17
61	7	89	74	51	8	5	3	4	19	20
62	3	62	62	63	5	5	0	4	17	17
63	17	93	78	88	10	5	5	3	19	17
64	12	63	79	63	9	5	4	2	17	11
65	2	78	66	77	10	5	5	3	23	25
66	4	84	43	83	10	5	5	5	20	23
67	16	52	36	78	7	5	2	1	13	15
68	3	91	60	78	10	5	5	5	21	22
69	6	80	97	45	7	5	2	4	24	26
70	21	95	94	85	9	4	5	4	15	14
71	10	72	27	57	4	4	0	3	22	25
72	6	65	64	64	7	5	2	3	10	17
73	16	72	53	68	10	5	5	5	17	17
74	21	84	54	58	5	5	0	2	20	21

ID	Section	ME 1	ME 2	ME 3	Pre MIE	Pre MIE (Non-Calc)	Pre MIE (Calc)	Post MIE (Calc)	Pre FCI	Post FCI
75	15	93	74	78	10	5	5	4	19	24
76	9	64	52	56	5	3	2	3	21	23
77	19	88	64	39	8	5	3	4	15	24
78	20	74	72	78	8	4	4	5	16	17
79	7	76	84	76	5	3	2	4	12	17
80	8	75	65	44	7	5	2	4	7	7
81	21	89	71	53	8	5	3	4	22	23
82	3	86	82	83	9	5	4	5	14	19
83	18	89	67	88	10	5	5	4	27	28
84	1	71	77	67	10	5	5	5	15	21
85	21	95	86	98	9	5	4	4	22	27
86	5	93	92	89	10	5	5	4	22	20
87	1	89	84	70	10	5	5	4	11	23
88	10	80	71	79	7	4	3	4	10	11
89	5	60	75	66	7	3	4	5	10	21
90	16	67	34	43	10	5	5	5	23	26
91	3	92	78	83	7	4	3	4	22	21
92	20	86	51	38	8	5	3	3	17	15
93	19	51	79	62	8	4	4	5	9	7
94	4	100	84	98	10	5	5	5	11	12
95	6	67	56	30	8	4	4	5	8	22
96	1	68	63	43	9	5	4	3	13	9
97	7	98	77	35	9	5	4	5	28	28
98	10	92	93	80	8	4	4	5	15	21
99	5	83	61	25	8	3	5	5	28	26
100	1	87	79	46	8	5	3	2	15	18
101	19	90	82	68	9	5	4	4	21	20
102	10	81	60	62	7	5	2	5	14	20
103	19	79	9	25	8	5	3	3	24	27
104	4	91	74	90	8	4	4	4	13	19
105	18	56	87	59	5	3	2	4	15	19
106	1	91	83	73	7	3	4	4	28	29
107	16	96	76	86	9	5	4	5	17	20
108	18	9	83	65	9	5	4	4	6	7
109	19	78	69	39	8	5	3	3	7	12
110	3	85	61	45	5	3	2	5	14	15
111	5	98	89	91	10	5	5	5	16	20
112	4	96	81	74	8	4	4	5	25	25
113	16	57	29	73	5	4	1	3	13	12

ID	Section	ME 1	ME 2	ME 3	Pre MIE	Pre MIE (Non-Calc)	Pre MIE (Calc)	Post MIE (Calc)	Pre FCI	Post FCI
114	2	71	29	81	7	5	2	2	8	11
115	6	98	91	57	10	5	5	5	18	23
116	12	58	57	45	8	5	3	2	11	13
117	10	79	76	42	7	4	3	4	7	14
118	21	95	93	83	7	3	4	4	26	27
119	9	90	92	85	7	5	2	5	15	22
120	10	86	85	81	6	3	3	4	13	23
121	3	91	82	76	9	5	4	3	26	27
122	7	86	67	84	5	3	2	5	10	26
123	12	89	76	75	7	3	4	4	22	26
124	12	70	67	55	6	3	3	4	9	8
125	5	95	87	87	9	4	5	5	18	16
126	11	69	86	68	6	5	1	4	11	14
127	6	81	83	56	7	4	3	5	8	22
128	21	75	73	68	4	2	2	3	11	10
129	3	74	89	51	10	5	5	5	16	27
130	1	89	59	35	9	5	4	3	23	11
131	5	98	89	94	9	5	4	4	22	25
132	1	63	92	75	10	5	5	5	9	8
133	10	93	93	80	8	5	3	5	17	23
134	4	85	81	77	8	5	3	4	16	16
135	13	91	79	78	7	3	4	3	18	22
136	9	78	96	83	9	5	4	4	16	22
137	10	98	88	59	9	5	4	4	11	12
138	3	96	84	74	10	5	5	5	11	17
139	2	94	74	63	7	5	2	4	13	24
140	1	74	74	56	9	5	4	4	9	8
141	18	51	72	68	8	4	4	4	12	19
142	2	64	95	100	9	5	4	4	24	26
143	21	96	98	64	8	5	3	4	23	23
144	18	82	70	44	10	5	5	3	11	11
145	19	89	83	50	8	4	4	5	11	23
146	2	97	67	93	7	4	3	3	16	20
147	11	48	89	55	9	5	4	5	18	27
148	21	93	82	71	9	5	4	4	21	25
149	6	100	87	102	10	5	5	5	28	28
150	2	47	47	58	10	5	5	4	18	27
151	14	86	85	66	8	4	4	3	17	23
152	5	87	80	51	8	4	4	5	12	20

ID	Section	ME 1	ME 2	ME 3	Pre MIE	Pre MIE (Non-Calc)	Pre MIE (Calc)	Post MIE (Calc)	Pre FCI	Post FCI
153	10	86	66	49	7	4	3	4	10	12
154	16	79	83	92	10	5	5	3	14	23
155	4	80	73	94	8	3	5	4	19	21
156	20	60	93	48	8	4	4	3	17	20
157	14	86	65	61	6	3	3	4	14	20
158	4	96	64	93	9	5	4	3	17	16
159	10	91	86	57	10	5	5	5	22	26
160	6	93	78	83	8	4	4	4	27	24
161	6	65	53	42	8	5	3	3	13	17
162	16	100	96	100	10	5	5	4	25	26
163	4	33	41	61	6	4	2	4	9	16
164	21	85	90	93	8	4	4	5	23	21
165	1	95	79	70	10	5	5	5	17	24
166	15	84	46	53	4	4	0	3	18	24
167	9	85	73	30	6	5	1	4	12	17
168	21	92	93	73	10	5	5	4	13	19
169	18	70	78	88	8	5	3	2	23	27
170	20	56	73	72	8	3	5	5	10	15
171	18	81	70	66	7	4	3	4	16	26
172	10	89	89	59	9	5	4	4	15	21
173	4	87	40	88	7	3	4	5	12	23
174	20	74	37	50	6	3	3	3	20	21
175	14	94	80	86	9	5	4	5	19	23
176	12	82	38	73	6	4	2	4	13	18
177	12	73	68	48	8	5	3	4	7	17
178	2	69	46	75	9	4	5	5	29	29
179	17	85	78	71	9	5	4	4	21	20
180	1	82	77	40	8	4	4	5	8	7
181	13	73	93	58	5	4	1	4	13	15
182	21	90	82	57	7	4	3	4	23	26
183	3	98	96	91	8	5	3	3	26	27
184	2	85	39	56	6	3	3	4	8	13
185	3	96	89	102	9	5	4	3	18	25
186	5	53	64	20	8	5	3	4	12	13
187	5	60	46	67	9	5	4	4	21	24
188	14	88	82	44	10	5	5	5	27	27
189	7	76	44	2	6	3	3	1	12	13
190	20	67	82	56	8	5	3	4	13	10
191	7	92	76	82	7	3	4	4	17	25

ID	Section	ME 1	ME 2	ME 3	Pre MIE	Pre MIE (Non-Calc)	Pre MIE (Calc)	Post MIE (Calc)	Pre FCI	Post FCI
192	10	83	86	78	8	5	3	5	22	24
193	5	95	82	77	8	4	4	4	15	19
194	15	84	79	27	6	5	1	4	19	24
195	12	76	53	52	8	4	4	2	14	13
196	20	92	53	65	10	5	5	5	20	26
197	19	60	84	57	8	4	4	5	10	16
198	16	95	64	88	8	5	3	4	23	22
199	19	40	57	27	4	3	1	1	4	13
200	13	69	88	50	7	3	4	4	9	16
201	2	64	70	75	10	5	5	3	11	13
202	18	78	77	95	9	5	4	4	12	19
203	2	43	88	87	9	5	4	4	26	26
204	16	70	51	50	9	5	4	5	8	10
205	21	91	91	70	8	5	3	4	23	26
206	6	98	90	97	9	4	5	4	17	17
207	20	80	81	73	8	4	4	5	19	20
208	12	93	63	48	10	5	5	5	23	23
209	6	51	77	51	4	4	0	4	9	11
210	20	48	78	50	7	3	4	4	11	13
211	12	68	59	39	8	4	4	4	10	13
212	16	89	65	91	9	5	4	4	6	10
213	14	70	56	63	6	4	2	2	21	26
214	20	86	86	65	7	4	3	5	26	25
215	16	93	71	90	7	4	3	5	22	25
216	21	98	88	79	10	5	5	5	28	28
217	17	98	71	82	9	5	4	4	19	24
218	5	86	64	56	10	5	5	5	10	26
219	4	43	48	59	7	2	5	5	13	12
220	20	72	71	45	9	5	4	5	9	15
221	3	79	87	73	7	4	3	4	6	16
222	19	66	75	58	8	5	3	2	6	16
223	6	66	48	42	8	4	4	4	10	16
224	4	73	47	71	7	4	3	3	10	18
225	2	84	69	71	7	4	3	5	23	20
226	16	63	33	58	7	5	2	5	4	7
227	18	66	60	34	9	4	5	4	9	11
228	3	83	51	81	9	4	5	3	23	26
229	2	66	27	54	6	4	2	3	25	24
230	12	90	78	84	9	5	4	3	20	24

ID	Section	ME 1	ME 2	ME 3	Pre MIE	Pre MIE (Non-Calc)	Pre MIE (Calc)	Post MIE (Calc)	Pre FCI	Post FCI
231	19	82	93	61	7	3	4	4	11	16
232	4	94	48	77	9	5	4	4	28	28
233	21	92	81	68	8	4	4	4	16	11
234	4	95	70	93	10	5	5	5	19	24
235	1	68	84	47	8	5	3	4	16	17
236	3	63	57	54	9	5	4	5	4	10
237	10	81	58	64	6	4	2	5	6	14
238	16	62	45	72	6	4	2	5	11	10
239	16	67	49	72	8	5	3	3	13	22
240	16	61	33	68	5	4	1	5	15	20
241	11	89	69	60	8	5	3	5	23	24
242	5	63	63	35	8	5	3	4	10	11
243	19	92	96	57	9	5	4	5	26	27
244	19	72	74	88	10	5	5	4	9	14
245	6	81	84	55	7	4	3	3	16	15
246	16	96	72	97	8	4	4	4	24	23
247	17	97	62	33	7	4	3	2	17	17
248	10	94	95	81	9	5	4	5	5	15
249	5	79	85	81	7	4	3	5	12	11
250	3	83	83	50	6	5	1	3	6	11
251	21	76	63	49	7	4	3	4	6	13
252	1	68	29	21	9	5	4	4	15	13
253	16	84	32	74	10	5	5	4	23	16
254	19	78	82	71	9	5	4	5	17	25
255	13	88	94	49	10	5	5	5	16	26
256	14	67	59	28	9	4	5	3	22	25
257	21	94	65	62	8	5	3	4	17	22
258	8	91	66	75	7	5	2	4	15	18
259	1	95	79	42	8	4	4	5	15	18
260	5	83	82	73	9	5	4	5	18	20
261	16	89	67	75	8	4	4	4	12	21
262	5	100	100	101	8	5	3	4	24	23
263	1	88	41	79	9	5	4	4	14	21
264	13	70	80	45	8	4	4	4	10	26
265	5	89	89	70	10	5	5	4	27	26
266	15	85	79	68	9	5	4	5	23	22
267	19	60	66	37	3	2	1	4	10	12
268	4	76	68	64	9	5	4	4	11	13
269	16	85	27	64	7	4	3	3	11	18

ID	Section	ME 1	ME 2	ME 3	Pre MIE	Pre MIE (Non-Calc)	Pre MIE (Calc)	Post MIE (Calc)	Pre FCI	Post FCI
270	3	75	73	72	10	5	5	4	18	20
271	6	77	80	62	8	5	3	3	4	5